MATHEMATICAL MODELING OF PHENYLALANINE AND LIGNIN BIOSYNTHETIC NETWORKS IN PLANTS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Longyun Guo

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. John A. Morgan, Chair

Davidson School of Chemical Engineering

Dr. Clint Chapple

Department of Biochemistry

Dr. Natalia Dudareva

Department of Biochemistry

Dr. Jennifer H. Wisecaver

Department of Biochemistry

Approved by:

Dr. Andrew D. Mesecar

Head of the School Graduate Program

ACKNOWLEDGMENTS

I would like to gratefully thank Dr. John A. Morgan at first for all his great guidance and supports during my graduate years. These helps are for sure invaluable since I could not have achieved current understanding of my research area without them. Throughout the years, he has given me one hundred percent trust, and has encouraged me to dig deeply into different kinds of mathematical ideas and techniques which might be more than sufficient for my ongoing project. With such freedom and resources, I am certain that my view of the world has greatly expanded and advanced, and the impacts will last much longer than any graduate program in the world. And when it comes specifically to my thesis research, He has made a great balance between the supervision and education. While I was free to explore all my ideas, he pointed out key direction for improvements when I struggled and got stuck. I am now proud of meeting all requirements of the graduation after years of learning, but I cannot imagine the accomplishment without him.

I would also like to thank my other thesis committee members, Dr. Clint Chapple, Dr. Natalia Dudareva, Dr. Jennifer H. Wisecaver, and Dr. David Rhodes. On the monthly group meetings for years, Dr. Clint Chapple and Dr. Natalia Dudareva have been both great tutors and audience for the project. Their discussions taught me how a scientist interprets the data and designs the experiments. Together with Dr. Jennifer H. Wisecaver and Dr. David Rhodes, their feedbacks for my presentations greatly improved my scientific communicating skills, especially to people from different area. Besides them, I also gratefully thank Dr. Jean Honorio for his excellent teaching in machine learning courses, as well as his great supports and guidance for me to explore new topics in mathematical theories.

Many thanks to my close collaborators Peng Wang and Rohit Jaini, who were working together with me to overcome all kinds of challenges emerging throughout the

project. Although sometimes it was frustrating to see things came out unexpectedly due to lack of experiences, I always felt inspired and energized again after analyzing all possible problems and solutions with them. Also as a modeler, I cannot be more grateful for their timely experimental supports. Even if it happens when experiments did not reproduce what model predicted (which seems like wasting their lives), they kept trusting me and we are together a great team.

Of course I will not forget and definitely appreciate all the helps from my other lab members and friends, especially Shaunak Ray. He has involved in the most of our project discussions, and provided different aspects of thinking by many valuable comments and questions. He is probably also the first applier of my published phenylpropanoid model, using some of my findings towards his own project. This connectivity of research makes me proud of what I have been done, letting me believe that my work could potentially have impacts to some other future research.

Finally, grateful love to my parents, who supported me without any doubts to pursue the doctoral degree at Purdue. There was a period that we three lived at three different countries over the world, which made me feel the planet is so large for us to stay close physically. Yet even just through the phones and videos, I saw their full presence by my side, which gave me enough inner peace to keep doing the things I like. This thesis is dedicated to them, Zhehao Guo and Mingji Cui.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

SYMBOLS

α	mRNA-protein tranfer coefficient
C	metabolite pool size
f	isotopic abundance
E	enzyme
G	gene
J	pathway flux
k	number of free parameters
K_a	activation constant
k_{cat}	turnover number
K_{decay}	mRNA decay coefficient
K_{ic}	competitive inhibition constant
K_{iu}	uncompetitive inhibition constant
K_m	binding affinity
K_{reg}	transcription factor regulatory coefficient
n	number of samples
P	protein
v	reaction velocity
V_{max}	maximal capacity
θ	unknown parameter

ABBREVIATIONS

4CL	4-coumarate: CoA ligase
5CAlc	5-hydroxyconiferyl alcohol
5CAld	5-hydroxyconiferaldehyde
5FA	5-hydroxyferulate
ADH	arogenate dehydrogenase
ADT	arogenate dehydratase
AIC	Akaike's information criterion
ALDH	aldehyde dehydrogenase
BLH	BELL-LIKE HOMEODOMAIN
C3'H	*p*-coumaroyl shikimate 3'-hydroxylase
C4H	cinnamate 4-hydroxylase
CA	cinnamate
CAlc	coniferyl alcohol
CAld	coniferaldehyde
CAD	cinnamyl alcohol dehydrogenase
CCoAOMT	caffeoyl CoA 3-O-methyltransferase
CCR	cinnamoyl-CoA reductase
CM	chorismate mutase
CoA	coenzyme A
COMT	caffeic acid O-methyltransferase
CSE	caffeoyl shikimate esterase
DAHP	3-deoxy-D-arabino-heptulosonate 7-phosphate
DREAM	dialogue for reverse engineering assessments and methods
E4P	erythrose 4-phosphate

F5H	ferulate 5-hydroxylase
FA	ferulate
FACoA	feruloyl CoA
FCC	flux control coefficient
FPKM	fragments per kilobase of transcript per million mapped reads
GRN	gene regulatory network
HCT	hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase
KA	caffeate
KACoA	caffeoyl CoA
KAlc	caffeoyl alcohol
KAld	caffealdehyde
KASK	caffeoyl shikimate
KFB	Kelch repeat F-box
KNAT	KNOTTED ARABIDOPSIS THALIANA
LC-MS	liquid chromatography-mass spectrometry
MCA	metabolic control analysis
MCMC	Markov Chain Monte Carlo
MED	Mediator complex
MFA	Metabolic flux analysis
MYB	myeloblastosis
NAC	NAM, ATAF1/2 and CUC2
NST	NAC SECONDARY WALL THICKENING PROMOTING FACTOR
PAL	phenylalanine ammonia lyase
*p*CA	*p*-coumarate
*p*CACoA	*p*-coumaroyl CoA
*p*CAlc	*p*-coumaryl alcohol
*p*CAld	*p*-coumaraldehyde
*p*CASK	*p*-coumaroyl shikimate

PCAT	plastidial cationic amino-acid transporter
PDT	prephenate dehydrogenase
PEP	phosphoenolpyruvate
PhCM2	petunia chorismate mutase 2
Phe	L-phenylalanine
PhpCAT	petunia plastidial cationic amino-acid transporter
PPA-AT	prephenate aminotransferase
PPY-AT	phenylpyruvate aminotransferase
PVT	putative vacuolar transporter
PXT	putative xylem transporter
SA	sinapate
SACoA	sinapoyl CoA
SAlc	sinapyl alcohol
SAld	sinapaldehyde
SND	SECONDARY WALL-ASSOCIATED NAC DOMIAN PROTEIN
SK	shikimate
VCAT	vacuolar cationic amino-acid transporter
VND	VASCULAR-RELATED NAC DOMAIN
TF	transcription factor
Trp	L-tryptophan
Tyr	L-tyrosine

ABSTRACT

Guo, Longyun PhD, Purdue University, May 2019. Mathematical Modeling of Phenylalanine and Lignin Biosynthetic Networks in Plants. Major Professor: John A. Morgan.

L-phenylalanine (Phe) is an important amino acid which is the precursor of various plant secondary metabolisms. Its biosynthesis and consumption are governed by different levels of regulatory mechanisms, yet our understanding to them are still far from complete. The plant has evolved a complex regulation over Phe, likely due to the fact that a significant portion of carbon assimilated by photosynthesis is diverted to its downstream products. In particular, lignin as one of them, is among the most abundant polymers in plant secondary cell wall. Studies have unraveled the interconnected metabolism involved in lignin biosynthesis, and a hierarchical gene regulatory network on top of it is also being uncovered by different research groups. These biological processes function together for sufficient lignification to ensure cell wall hydrophobicity and rigidity for plant normal growth. Yet on the other hand, the presence of lignin hinders the efficient saccharification process for biofuel production. Therefore, it is fundamental to understand lignin biosynthesis and its upstream Phe biosynthesis in a systematic way, to guide rational metabolic engineering to either reduce lignin content or manipulate its composition *in planta*.

Phe biosynthesis was predominantly existed in plastids according to previous studies, and there exists a cytosolic synthetic route as well. Yet how two pathways are metabolically coordinated are largely under-explored. Here I describe a flux analysis using time course datasets from ^{15}N L-tyrosine (Tyr) isotopic labeling studies to show the contributions from two alternative Phe biosynthetic routes in Petunia flower. The flux split between cytosolic and plastidial routes were sensitive to genetic

perturbations to either upstream chorismate mutase within shikimate pathway, or downstream plastidial cationic amino-acid transporter. These results indicate the biological significance of having an alternative biosynthetic route to this important amino acid, so that defects of the plastidial route can be partially compensated to maintain Phe homeostasis.

To understand the metabolic dynamics of the upstream part of lignin biosynthesis, we developed a multicompartmental kinetic model of the general phenylpropanoid metabolism in Arabidopsis basal lignifying stems. The model was parameterized by Markov Chain Monte Carlo sampling, with data from feeding plants with ring labeled [$^{13}C_6$]-Phe. The existence of vacuole storage for both Phe and *p*-coumarate was supported by an information theoretic approach. Metabolic control analysis with the model suggested the plastidial cationic amino-acid transporter to be the step with the highest flux controlling coefficient for lignin deposition rate. This model provides a deeper understanding of the metabolic connections between Phe biosynthesis and phenylpropanoid metabolism, suggesting the transporter step to be the promising target if one aims to manipulate lignin pathway flux.

Hundreds of gene regulatory interactions between transcription factors and structural genes involved in lignin biosynthesis has been reported with different experimental evidence in model plant Arabidopsis, however, a public database is missing to summarize and present all these findings. In this work, we documented all reported gene regulatory interactions in Arabidopsis lignin biosynthesis, and ended up with a gene regulatory network consisting of 438 interactions between 72 genes. A network is then constructed with linear differential equations, and its parameters were estimated and evaluated with RNA-seq datasets from 13 genetic backgrounds in Arabidopsis basal stems. We combined this network with a kinetic model of lignin biosynthesis starting from Phe and ending with all monolignols participated in lignin polymerization. This hierarchical kinetic model is the first model integrating dynamic information between transcriptional machinery and metabolic network for lignin biosynthesis. We showed that it is able to provide mechanistic explanations for most of experimental findings

from different genotypes. It also provides the opportunity to systematically test all possible genetic manipulation strategies targeting to lignification relevant genes to predict the lignin phenotypes *in silico*.

1. INTRODUCTION

1.1 Phenylalanine biosynthetic network in plants

L-phenylalanine (Phe) is one of nine essential amino acids which is not produced in the human body. Together with two other aromatic amino acids, namely L-tyrosine (Tyr) and L-tryptophan (Trp), Phe is synthesized by the shikimate pathway in plants [1]. Thousands of phenylalanine-derived metabolites have been identified, and they play important roles in different aspects for plants normal growth under various environmental conditions. These metabolites include phenylpropanoids, benzenoids, salicylic acids, flavonoids, etc. Many of these are of essential for normal plant development. For instance, the biopolymer lignin is one of major secondary cell wall components, and its subunits are all derived from Phe. Thus not surprisingly, a significant amount of carbon fixed by plants is invested into Phe biosynthesis [2].

Phe is known to be synthesized via two alternative pathways from prephenate, namely, the arogenate pathway and phenylpyruvate pathway [3]. In the arogenate pathway, prephenate is first transaminated and then dehydrated/decarboxylated to phenylalanine by prephenate aminotransferase (PPA-AT) and arogenate dehydratase (ADT). On the other hand, in the phenylpyruvate pathway, Phe is produced with prephenate dehydratase (PDT) followed by phenylpyruvate aminotransferase (PPY-AT). [4] have reported that Phe is predominantly produced in plastids via the arogenate pathway in Petunia. The dominance of the plastidial route for Phe biosynthesis was further verified in later studies in tomato [5] and Arabidopsis [6]. On the other hand, [7] identified a cytosolic phenylpyruvate pathway which alternatively produces Phe in Petunia. Interestingly, cytosolic PPY-AT strongly favors Tyr as the amino donor [7] instead of glutamate. The alternative cytosolic route seems to play a compensating role for Phe homeostasis, as its flux was increased when ADT and PPA-AT were

transiently downregulated in petunia flowers [7]. These studies together indicate that a complex Phe biosynthetic network spanning multiple compartments is present in plants.

As an important precursor to various downstream metabolisms, Phe biosynthesis is regulated by different regulatory mechanisms in plants [1]. Multiple allosteric regulatory mechanisms are reported for Phe biosynthesis in plants. Phe was shown to competitively inhibit chorismate mutase (CM) [8, 9] and ADT [10, 11]. While there is no evidence for a Phe-sensitive 3-Deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase in plants, inhibition of DAHP synthase from Tyr and Trp were reported [12, 13]. Furthermore, the fact that Tyr activates ADT [10] and Trp activates CM [8, 9] suggests the complex interactions exist among the aromatic amino acids within shikimate pathway. Beyond the allosteric regulation, a transcription factor ODORANT1 was identified in Petunia flowers and its suppression through RNAi led to decreased expression for genes encoding DAHP synthase and 5-enol-pyruvyl-shikimate-3-phosphate synthase, which are involved in shikimate pathway [14]. [4] later found the increased expression of the shikimate pathway genes with reduced Phe levels when ADT was suppressed by RNAi. These results suggest that in addition to allosteric regulation, transcriptional regulation connected with a currently unknown signaling mechanism might be also active in Phe biosynthetic network in plants.

1.2 Lignin biosynthesis in plants

Lignin is a heteropolymer consisting primarily of subunits which are all derived from Phe. As one of the key components in terrestrial plant secondary cell wall, lignin provides structural supports for plants and it also enables efficient water transport by its hydrophobicity [15]. On the other hand, lignin impedes the enzymatic hydrolysis of cell wall polysaccharides for biofuel production, by physically detaching polysaccharides from the hydrolyzing enzyme or competitively inhibiting the reaction [16]. Genetic manipulation of lignin synthesis is thus one potential approach to lower

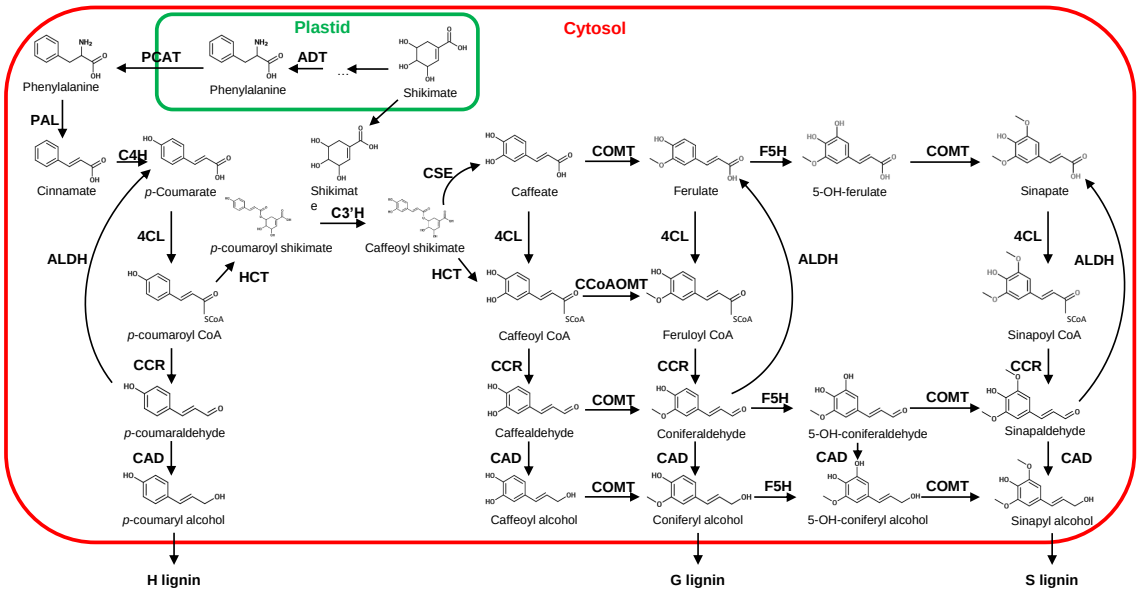

Fig. 1.1. Lignin biosynthesis in Arabidopsis. Abbreviations: 4CL, 4-coumarate: CoA ligase; ADT, arogenate dehydratase; ALDH, aldehyde dehydrogenase; C3H, *p*-coumaroyl shikimate 3-hydroxylase; C4H, cinnamate-4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCoAOMT, caffeoyl CoA 3-O-methyltransferase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid O-methyltransferase; CSE, caffeoyl shikimate esterase; F5H, ferulate 5-hydroxylase; HCT, hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase; PAL, phenylalanine ammonia lyase; PCAT, plastidial cationic amino-acid transporter.

the pretreatment costs, by either reducing the total lignin content or altering its composition [17]. While success has been achieved by perturbating the expression of genes involved in lignin biosynthesis [18–23], undesired phenotypes such as dwarfism were observed in some cases [24–28], indicating lignification is a complex process of which the perturbation responses cannot be easily predicted.

In plants, lignin is mainly composed of three monolignols, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol. All three alcohols are synthesized from cytosolic Phe, and twelve enzyme families are responsible for monolignol biosynthesis in

Arabidopsis (Fig. 1.1. Isoforms exist for 4-coumarate: CoA ligases (4CL) [29–31], cinnamyl alcohol dehydrogenases (CAD) [32, 33] and phenylalanine ammonia lyases (PAL) [34,35], which are known to be directly involved in Arabidopsis lignin biosynthesis. Some other enzymes have isoforms active in response to environmental stimuli, for instance, the Arabidopsis gene for cinnamoyl-CoA reductase 2 (CCR2) was shown to be upregulated under biotic stress [36]. Further complexities are introduced with some enzymes able to catalyze multiple substrates (Fig. 1.1. In Arabidopsis, 4CLs show activities for *p*-coumarate, caffeate, ferulate and sinapate [29,30]; all CoA esters in the pathway are reduced by CCR1 [36, 37]; Aldehyde dehydrogenase (ALDH) can utilize *p*-coumaraldehyde, caffealdehyde and sinapaldehyde [38]; both CAD4 and CAD5 can reduce all aldehydes in the pathway [33]; Hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT) is responsible for two reversible reactions between CoA esters and shikimate esters [39]; Ferulate 5-hydroxylase (F5H) has activities for coniferaldehyde and coniferyl alcohol, and marginally for ferulate as well [40]; Caffeic acid O-methyltransferase (COMT) attaches a methyl group to a wide range of pathway intermediates [41]. Together they form a complicated network for lignin biosynthesis (Fig. 1.1, and modeling efforts are definitely needed to systematically understand the individual roles of enzymes towards overall pathway dynamics.

To ensure sufficient lignin deposition during the plant normal development, various transcription factors function to control the expression of structural genes in lignin biosynthesis, usually in coordination with genes responsible for other secondary cell wall components [42, 43]. Multiple NAM, ATAF1/2 and CUC2 (NAC) master switches have been identified to activate the second-level myeloblastosis (MYB) factors, and they together induce the expression of other downstream transcription factors and secondary cell wall biosynthetic genes [44–50]. While it is believed that a well-structured hierarchy exists for gene regulatory network of lignin biosynthesis [51], the real situation might be more complicated, as some MYB factors can affect the expression of NAC genes, forming feedback loops within the gene regulatory network. For instances in Arabidopsis, MYB4, MYB7 and MYB32 repressed the expression of SECONDARY

WALL-ASSOCIATED NAC DOMAIN PROTEIN 1 (SND1) in transfection assays, while SND1 was shown to directly bind to MYB32 promoter region in an electrophoretic mobility shift assay [52]. MYB4, MYB7 and MYB32 were all directly activated by MYB46 [53, 54], and MYB46 can be directly bound and activated by SND1 [53, 55, 56]. Interestingly, the expression of MYB7 can be repressed by MYB4 [57]. On the other hand, the expression of MYB4 is upregulated by subunits of Mediator complex, MED5a and MED5b [58]. In addition, the Mediator complex is involved in the repression of gene expression in lignin biosynthesis, probably by regulating MYB factors including MYB4, MYB20, MYB43, MYB85 and MYB103 [58]. The gene expression of Kelch repeat F-box (KFB) protein family is also regulated by Mediator [58], and KFBs are known to mediate PAL activities by ubiquitination followed by degradation [59–61]. Why plants have evolved such a complex gene regulatory network is still unclear, and a modeling approach integrated with transcriptional analysis will help to organize the current knowledge of gene-gene interactions in Arabidopsis, as well as generating new hypotheses to test for a better understanding of the regulation of lignin biosynthesis.

1.3 Mathematical approaches to understand cellular dynamics in biological systems

Various experimental techniques are currently widely available to quantify different components in a biological system. For instances, RNA-seq can be used to measure the gene expressions in a genome scale [62]. Proteomic and metabolomic data can be obtained with MS based approaches [63–65]. However, no techniques are available to directly quantify intracellular metabolic fluxes. This is mainly due to the fact that metabolic fluxes are an emergent property of the biological system, resulting from the interactions between enzymes and intermediates. If the studied pathway is strictly linear and its product is exported extracellularly, internal fluxes can then be recovered from the excretion measurements. However, such simplicity is rare because multiple metabolic branch points are usually present to either ensure the needs from different

downstream metabolisms or due to the enzyme promiscuity (e.g., 4-coumaric acid: coenzyme A lyases are able to use *p*-coumarate, caffeate, ferulate, as well as sinapate as substrates in lignin biosynthesis). Therefore, metabolic fluxes are usually indirectly estimated with metabolic flux analysis which fits data from isotopic labeling studies using metabolite and isotopic atom enrichment balance equations. On the other hand, kinetic modeling can also reveal the flux distribution of the target metabolism, as it directly simulates the interactions between enzymes and metabolites within the pathway. This property makes the applications of the kinetic model wider than metabolic flux analysis, as it can be used to simulate the consequences when certain parts of the metabolism are perturbed. Next, these techniques will be described in more detail and their applications towards the Phe and lignin biosynthetic network in plants are presented in the current dissertation.

1.3.1 Metabolic flux analysis

Metabolic flux analysis (MFA) is widely applied in different studies to quantify the metabolic fluxes of the target metabolism [66, 67]. Frequently, an isotopically labeled precursor is fed into the system, and metabolic flow through pathway reactions can be indirectly quantified with the isotope enrichment of the intermediate pools. Depending on whether the measurements are taken at an isotopically steady state or in a time series before the steady state, MFA is applied as either stationary MFA [68] or a non-stationary MFA [69, 70]. In either case, mathematical modeling is needed to recover the flux information from the isotope enrichment data. There are two kinds of balance relationships in a typical MFA approach; one for the balance between the synthesis and the consumption for each metabolite, which is known as the mass balance. The other is an enrichment balance, as its name suggested, describing the

fractions of isotopic and natural abundances for each metabolite. Both balances are represented as differential equations listed below:

$$\frac{dC}{dt} = \sum_{i}^{synthesis} v_i - \sum_{j}^{consumption} v_j \tag{1.1}$$

$$\frac{dfC}{dt} = \sum_{i}^{synthesis} f_i v_i - \sum_{j}^{consumption} f v_j \tag{1.2}$$

where t is time, C is the concentration of the metabolite, f is the isotopic enrichment of the metabolite, and v is the velocity of the relevant reaction.

A metabolic steady state is usually assumed for the biological system, meaning all fluxes are constant over time. This assumption avoids the needs to simulate enzyme kinetics, thus greatly simplifies the model structure. In this case, mass balances are reduced to a set of linear equations to constrain the flux values. This leads to a smaller set of independent fluxes, and the rest of dependent fluxes can be readily deduced with the linear equations once the independent fluxes are quantified. For a non-stationary MFA, isotopic dynamics from measurements are simulated with the adjusted isotopic enrichment balances:

$$\frac{df}{dt} = \frac{\sum_i^{synthesis} f_i v_i - \sum_j^{consumption} f v_j}{C} \tag{1.3}$$

This is obtained with $\frac{dC}{dt}$ being zero due to a metabolic steady state. An optimization algorithm is then applied to estimate all independent fluxes and metabolite concentrations, so that the isotopic enrichment dynamics derived from eq. 1.3 match with experimental observations.

It is simpler to perform MFA in single celled microbial systems, as they are closer to a homogenous system than multicompartmental organisms. This means the isotopically labeled molecules are well mixed with the unlabeled cellular intermediates, so that the pathway dynamics can be reliably estimated with eq. 1.3. Some applications of MFA in microbes include the studies of central carbon metabolism in cyanobacteria [69], *Escherichia coli* [71] and a thermophilic strain *Geobacillus* LC300 [72], as well as a genome-scale MFA in *Escherichia coli* [73]. In plants, additional parameters are

needed to quantify the subcellular pool sizes for multicompartmental metabolites. [74] have applied MFA to generate flux maps for photosynthetic metabolism in Arabidopsis under varying light conditions. They found that photorespiration was enhanced under high light condition in coordination with higher CO_2 fixation. This is probably due to the lower CO_2 partial pressure at the carboxylation site in high light acclimated plants [75]. In another example of plant MFA, [76] have determined the flux distribution for oilseed rape embryos of nine genotypes with different biomass compositions. They observed a tradeoff between the contents of lipid and starch, and this tradeoff is likely mediated by post-translational regulation of phosphofructokinase and ADP-glucose pyrophosphorylase. MFA has been also applied to study the cellular dynamics under low-nitrogen condition for cell suspensions of hybrid poplar [77]. Other examples of MFA in plants can be found in a recent review [78]. In certain cases, the assumption of a metabolic steady state has to be relaxed if the whole system obviously obviates from the steady state over time. In [79], [80] and [81], we assumed the rate of change of fluxes was linear over time, which ended up with a similar form of MFA allowing one to reliably solve for the fluxes from the corresponding isotopic labeling studies.

1.3.2 Kinetic modeling

Kinetic modeling is a mathematical approach to describe cellular metabolism in a mechanistic manner. Most of its applications in plants are focused on a target pathway in a specified tissue [82, 83]; Fig. 1.2 and Supplementary Table A.1. Similar with MFA, it uses mass balance eq. 1.1 to describe the metabolic dynamics within a pathway. In contrast to MFA, the reaction velocity v is no longer represented by a fixed parameter, but it is simulated as a rate equation, which is a function of the concentrations of substrates, the catalyzing enzyme, as well as other metabolites as allosteric regulators if necessary. Michaelis-Menten kinetics are usually the default choices for the rate equations [84,85]. However, other kinds of functions are also widely used in different kinetic models, including Hill equation [86, 87], mass action [88], and

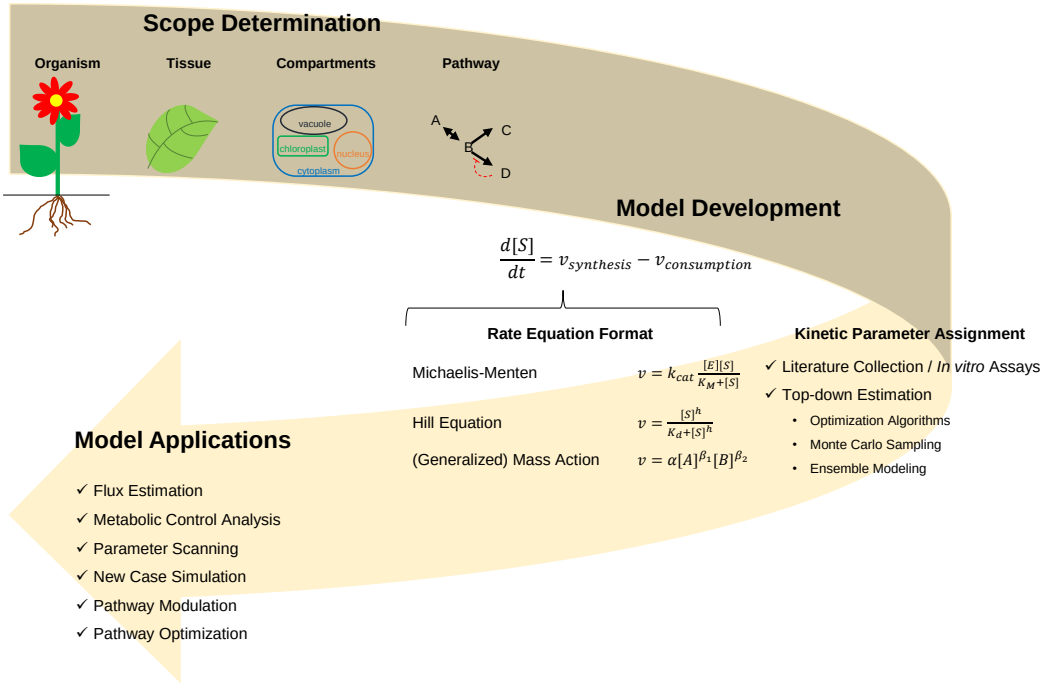

Fig. 1.2. Workflow of kinetic modeling in plants.

generalized mass action [89, 90] (Fig. 1.2 and Supplementary Table A.1. Specifically, the Hill equation is used when substrate cooperativity is observed for the corresponding enzyme kinetics [91]. And if no mechanistic kinetic information is available for the reactions, generalized mass action is a good choice to empirically fit the data [92–95].

If most of enzymes within the target pathway are already characterized with in vitro assays, the easiest way to parameterize the kinetic model is to apply those kinetic parameters to the model [91, 96–98]. Since enzyme amounts vary between different species and tissues, they need to be quantified from the studied organism instead of any in vitro systems. In plants, this approach has been taken to study sucrose metabolism [96], photosynthesis [97, 98], aspartate-derived amino-acid pathway [91], etc. It should be noted that care must be taken when one implements these models, as the *in vitro* assays conditions might be distinct from the physiological one [99]. The

best practice is thus performing any needed *in vitro* assays under conditions that best mimic the cellular environment. Ideally, kinetic models built in this way are shown to have good agreement with the *in vivo* dynamics [91]. While this bottom-up approach is straightforward, a top-down parameter estimation is necessary when some or all of the parameters are missing. Indeed, the situation of lacking literature values for enzyme parameter references is frequently encountered, especially when it comes to secondary metabolism [100–103]. In this case, this top-down approach is essentially an optimization problem, where the objective function is to minimize the differences between the model predicted dynamics and available measurements by adjusting the values of free parameters. Various optimization algorithms were applied to solve the parameter estimation problem [101, 103–105], yet because of the non-convexity of the problem, none of them guarantee to identify the global minimum. This limitation makes the models fidelity questionable, and careful validation with independent datasets is indispensable to verify that the model trained in this way correctly reflects *in vivo* behaviors. Some successful examples using this optimization framework include the parameterization of phenylpropanoid metabolism [100], fenclorim metabolism [101], flavonoid pathway [102] and benzenoid network [103].

There are other methodologies for parameter estimation, which partially overcome the limitation of the current optimization algorithms. Ensemble modeling increases the reliability of the modeling output by summarizing outputs from thousands or even millions of individual models [92–95, 106]. Each individual model is parameterized with a sample from a prior parameter distribution, which is usually a uniform distribution spanning the possible physiological range. One model is selected into the ensemble only when it passes the pre-specified criteria, which is the minimum agreement it needs to match with the known facts. Therefore, although for each model, it might fit the data up to a suboptimal degree, the sum of many of them could end up with an output range covering the reality with a high confidence. The limitation of ensemble modeling is that the choices of sample number and filtering criteria are somewhat arbitrary. Moreover, when the parameter distribution matching the observation is significantly

narrower than the prior distribution, the valid samples are very difficult to obtain. A series of models for lignin biosynthesis in different plant species were developed using ensemble modeling [92–95], and hypotheses of metabolic channeling by multiple enzymes were proposed based on comparison between models with different structure.

Monte Carlo sampling is another methodology to infer parameter values by estimating the distribution of parameters given data [107]. We implemented Markov Chain Monte Carlo (MCMC) sampling in our parameter estimation approach [108, 109]. The advantage of MCMC is that it guarantees to reach the global minimum given sufficient number of samples. Briefly, in a typical MCMC approach, a Markov Chain of samples is generated based on a pre-specified transition function. The transition function is composed of a proposal function and an acceptance function:

$$T(x \to x') = q(x'|x)\alpha(x', x) \tag{1.4}$$

In Metropolis-Hastings algorithm, the proposal function is a symmetric one (See eq. 1.5 for an instance of a Gaussian random walk with a step size of σ), and the acceptance function is calculated based on the likelihood ratio of the proposed sample x' and the previous sample x (eq. 1.6 [110, 111].

$$x' \sim q(x'|x) \equiv N(x, \sigma^2) \tag{1.5}$$

$$\alpha(x', x) = min\{1, \frac{\pi(x')}{\pi(x)}\} \tag{1.6}$$

$$\pi(x)T(x \to x') = \pi(x')T(x' \to x) \tag{1.7}$$

In this setting, the detailed balance of a Markov Chain (eq. 1.7 is preserved, so that given a sufficiently long run, the sample distribution generated by the transition function will converge to the posterior distribution π due to the property of a Markov Chain.

One key point for a successful MCMC is to control the balance between the acceptance rate and the proposing step size. It is usually the case that when a step size is too large, the acceptance rate will be very low since the new sample is likely picked in a low probability area. On the other hand, although a small step size increases the

acceptance rate, it could get trapped into a local probability area thus inefficiently exploring the whole parameter space [107]. While any proposal functions can be used as long as they satisfy certain criteria, a proposal with bad mixing property can lead to a very slow convergence. An advanced version of MCMC has been proposed in [108], which greatly increases the MCMC mixing property, so that it could shift to a higher probability area faster. However, when it comes to a high dimensional problem, even this advanced MCMC could suffer from the extremely high computation demands for parameter estimation.

Once the kinetic model is developed, it can be used in different ways. One of the most common applications is metabolic control analysis [112, 113]. As is suggested by its name, metabolic control analysis is performed to calculate the changes of the target flux when each network component is slightly perturbed. For an instance, flux control coefficient for individual enzymes within the pathway can be obtained with this analysis, and it has been shown that instead of only single enzyme being the sole limiting step determining the pathway flux, the control of flux is spread to multiple enzymes in most cases. Other applications of kinetic models include: flux estimation; simulation of metabolic responses to genetic or environmental changes; parameter scanning, where one adjusts the value of a given parameter in a specified range to see the models sensitivity to the parameter [114, 115]; pathway modulation, where one monitors the significance of each kinetic mechanism by removing it from the model and observe the effect [109, 116]; and pathway optimization, where one simulates the optimal adjustments to the pathway to achieve a given objective, and the objective can be a metabolic engineering goal [92, 97, 98].

1.3.3 Gene regulatory network inference

The transcriptional rate of a given gene is largely controlled by its upstream transcription factors or other regulators. By defining a directional relationship between the upstream transcription factor and its downstream target gene, a gene regulatory

network (GRN) is then collectively defined by these interactions. It is well known that complex GRNs exist, yet a direct test of all possible transcription factor and gene interactions in a genome is very difficult if not impossible. Given the relative easier measurements of mRNA levels, many attempts have been made to infer a GRN given expression datasets using all kinds of mathematical techniques, which are summarized in recent reviews [117–119]. These techniques include regression-based methods [120–123], Bayesian approach [124, 125], differential equations [126–131], as well as other kinds of statistical or machine learning methods [132–136]. The generated GRN can be either semi-quantitative, with only interactions recovered from the data [121, 134, 135]; or quantitatively differentiate interactions based on their regulatory strengths, therefore can be used to simulate cases when certain genes are silenced within the network through genetic engineering [125–128, 131].

To date, many newly developed GRN inference methods were tested for their efficiencies on the Dialogue for Reverse Engineering Assessments and Methods (DREAM) challenge benchmarks. For instances, a tree-based ensemble method GENIE3 [134] ranked the first among the algorithms that took part in DREAM4 competition [137], where gene-gene interactions were inferred from the synthetic gene expression datasets generated in silico by networks of up to 100 genes. GENIE3 continued to win the DREAM5 competition [138], closely followed by an ANOVA based method ANOVerence [135]. It is worth mentioning that DREAM5 consists of microarray datasets from *Escherichia coli*, *Staphylococcus aureus* and *Saccharomyces*, each under hundreds of conditions generated by different gene, drug and environmental perturbations [138]. A later published method, regression-based TIGRESS [121], was shown to have similar performance as GENIE3 and ANOVerence, and its performance can be significantly improved by a fine parameter tuning. A recent GRN inference method based on Granger causality [136] was verified with DREAM3, a benchmark similar to DREAM4 [139]. Since each of these methods are developed with different mathematical foundations, it is a common approach to average their predictions for a robust GRN inference in practice. [140] generated an Arabidopsis GRN for

secondary cell wall synthesis using GENIE3, ANOVerence, TIGRESS, Inferelator [133] and SIRENE [132]. Other applications of GRN inference in plants include models for the circadian clock [126, 129], flowering time [130], seed development [141], fruit development [142], root hair growth [143], etc.

One advantage of using differential equations to reconstruct a GRN is their compatibility with metabolic dynamic models [129]. Although theoretically sound, current models integrating transcriptional and metabolic levels are mainly applying simpler settings for each level [144]. Flux balance analysis is heavily applied for metabolism, and transcriptional regulation is integrated with techniques such as Boolean logic or some statistical methods [145–147]. This is almost certainly due to the difficulties in parameterization of a large differential equation system. In the current study, we partially solve this challenge by selecting a candidate interaction set *a priori* from literature evidence, and by assuming linearity for the gene regulatory interactions. A GRN consisting of 250 interactions is integrated with a kinetic model of lignin biosynthesis having 50 metabolic reactions in total.

1.4 Organization of dissertation

The dissertation is organized as follows. In chapter 2, flux analysis of the Phe biosynthetic network in plants is presented. This analysis directly extracted the fluxes from several isotopic labeling studies, and avoids the need of detailed enzyme kinetics for each metabolic step. Interplay between plastidial arogenate pathway and cytosolic phenylpyruvate pathway was elucidated, as well as several regulatory mechanisms for Phe homeostasis. Some of these key findings were incorporated into the kinetic model of phenylpropanoid metabolism in Arabidopsis, which is described in chapter 3. An information theoretic approach helped to select the best model from multiple candidates, and revealed the flux controlling property of the Phe export step from the plastid. Such understanding of the upstream lignin biosynthesis was the initial building block of an extended kinetic model covering the whole lignin metabolism, including a

gene regulatory network on top of lignification in Arabidopsis. The construction and the application of this hierarchical model was presented in chapter 4.

1.5 Statement of published and collaborative work

Chapter 2 contains materials published in Nature Communications, Volume 6, 2015 under doi 10.1038/ncomms9142; Plant Journal, Volume 92, 2017 under doi 10.1111/tpj.13730 and Nature Communications, Volume 10, 2019 under doi 10.1038/s41467-018-07969-2. I performed the mathematical modeling for all published work and wrote the Methods for computation parts.

Chapter 3 has been published in Metabolic Engineering, Volume 49, 2018 under doi 10.1016/j.ymben.2018.07.003. I performed the mathematical modeling of the published work and wrote the main body of the manuscript. Peng Wang performed the feeding experiments and contributed equally as co-first author for the manuscript.

Materials in Chapter 1 and Chapter 4 will be published in near future.

2. FLUX ANALYSIS UNCOVERING KEY REGULATORY MECHANISMS IN PHENYLALANINE METABOLISM

2.1 Introduction

In addition to be involved in the protein biosynthesis, L-phenylalanine (Phe) is also the precursor of thousands of indispensable metabolites in plants [1]. These phenolic compounds play vital roles in various processes to ensure plants normal growth, reproduction and survival. For example, the biopolymer lignin is derived from Phe and functions to sustain rigidity and hydrophobicity of cell walls to provide enough stem strength and efficient water transduction [17]. Other important Phe-derived metabolites include phenylpropanoids, benzenoids, salicylic acids, flavonoids, etc. While most of these compounds are synthesized in the cytosol, their precursor Phe is predominantly produced in plastids. Inside the plastid, the shikimate pathway is the precursor to produce all three aromatic amino acids, Phe, L-tyrosine (Tyr) and L-tryptophan (Trp) [1]. Complex regulatory mechanisms are active in this pathway to accurately control the carbon partition towards each of the aromatic amino acids, which consist of several allosteric interactions between end products and upstream catalyzing enzymes [12, 13, 148, 149]. Transcriptional regulation is also present as the expression of shikimate pathway genes is induced upon various environmental stimuli [150,151]. However, whether there exist any transcriptional mechanisms sensing downstream metabolite concentrations is still unclear.

A protein transporter is required for Phe translocation towards the cytosol, due to Phes low permeability through the membrane [152, 153]. Our collaborator has identified a gene encoding a Petunia hybrida plastidial cationic amino-acid transporter (PhpCAT) located in plastids [79]. This transporter is capable of transporting all three aromatic amino acids, and its key role in Phe metabolism was characterized

with a series of molecular biological and genetic studies [79]. Thus [79] has completed the final metabolic step connecting Phe biosynthesis in plastids and Phe consumption in cytosol.

Previously, [7] identified an alternative Phe biosynthetic route in cytosol, which produces Phe with a cytosolic phenylpyruvate aminotransferase (PPY-AT) instead of arogenate dehydratase (ADT) (Fig. 2.1a). While PPY-AT is shown to be actively contributing to make Phe, it is unknown about the starting point of this cytosolic pathway. Recently, [81] has revisited the cytosolic isoform PhCM2 in Petunia, and demonstrated its role in providing the precursor prephenate for the downstream prephenate dehydratase (PDT) and following PPY-AT for Phe biosynthesis. RNAi suppression of *PhCM2* has led to the decreased levels of all three aromatic amino acids, suggesting some yet unknown regulatory mechanisms on top of the two alternative pathways.

Plants dont have similar Phe catabolic enzymes as mammals and microbes [154,155]. Hyperaccumulation of Phe was observed when its consumption through phenylalanine ammonia lyase (PAL) was impaired in Arabidopsis *pal1 pal2* lines [35]. These observations raise the question that how plants deal with high Phe levels. *PAL*-RNAi lines were generated in Petunia and all three aromatic amino acids were significantly accumulated compared to wild type (Fig. 1 in [80]. Interestingly, shikimate concentrations in *PAL*-RNAi lines were largely reduced, indicating the suppression of shikimate pathway [80]. A *Petunia hybrida* tonoplast cationic amino-acid transporter (PhCAT2) was identified and its role in the vacuolar sequestration of Phe was further characterized [80]. The presence of PhCAT2 explains a nearly twofold expansion in Phe content in *PAL*-RNAi lines relative to wild type vacuoles. The reduction of its activity in *PAL*-RNAi lines led to higher phenylacetaldehyde emission, which was likely due to the increased Phe availability in cytosol [80].

Here we used flowers of *Petunia hybrida* cv Mitchell, a plant system with high carbon flux towards emitted phenylalanine-derived volatiles, to study the Phe metabolism by flux modeling strategies without the need of enzyme kinetic information. These models

are particularly suitable for the case where full knowledge of the enzymatic steps is unavailable, while metabolic flux values can still be reliably estimated using data from isotopic labeling studies. Carbon flux through the cytosolic phenylpyruvate pathway was tracked with ^{15}N-tyrosine, as the cytosolic aminotransferase has a high affinity for tyrosine as the amino donor. Using data from this isotopic study, we successfully revealed the flux dynamics of two alternative pathways for Phe biosynthesis between 18:00 and 22:00 h for 2-day-old Petunia corollas. In addition, further studies are needed to investigate whether the perturbation of cytosolic Phe consumpton could affect its upstream plastidial shikimate pathway. To keep track of the flux splits between the aromatic amino acids from shikimate, uniformly ^{13}C-labeled ([U-$^{13}C_6$]) glucose was applied to the flower system. A different flux model with this dataset was developed to show the different flux distributions between wild-type and *PAL*-RNAi lines in Petunia.

2.2 Materials and Methods

All simulations were performed in Matlab R2013a environment (The MathWorks, Inc., Natick, MA).

2.2.1 Metabolic flux analysis with ^{15}N-tyrosine labeling

We generated a metabolic flux model that utilizes experimentally determined pool sizes and isotopic enrichment of phenylalanine from exogenously fed ^{15}N-tyrosine. Feeding of 10 mM ^{15}N-tyrosine (Cambridge Isotope Laboratories, Andover, MA) to 2-day-old corollas of control and *PhpCAT*-RNAi lines was performed as previously described [7]. Similar to our previous labeling study [7], the labeling percentage of tyrosine quickly reached >80% and a pseudo-steady state within 2 h upon feeding,

and stayed constant for the rest of the experimental period. The model was based on the dynamic mass balance around phenylalanine as defined as:

$$\frac{dC_{Phe}}{dt} = v_1 + v_2 - v_c \tag{2.1}$$

where C_{Phe} is the phenylalanine pool size, v_1 is the flux through the plastidial arogenate pathway, v_2 is the flux through the cytosolic phenylpyruvate pathway and v_c is the consumption rate of phenylalanine. To determine v_2, it was also required to know the mass balance of isotopic enrichment for phenylalanine as defined as:

$$\frac{df_{Phe}C_{Phe}}{dt} = f_{Tyr}v_2 - f_{Phe}v_c \tag{2.2}$$

where f_{Phe} and f_{Tyr} represent isotopic abundance of phenylalanine and tyrosine, respectively, in the total pools. With equations 2.1 and 2.2, v_1 and v_2 can be derived as follows:

$$v_1 = \frac{(f_{Tyr} - f_{Phe})\frac{dC_{Phe}}{dt}C_{Phe}\frac{df_{Phe}}{dt}}{f_{Tyr}} + \frac{f_{Tyr} - f_{Phe}}{f_{Tyr}}v_c \tag{2.3}$$

$$v_2 = \frac{f_{Phe}\frac{dC_{Phe}}{dt} + C_{Phe}\frac{df_{Phe}}{dt}}{f_{Tyr}} + \frac{f_{Phe}}{f_{Tyr}}v_c \tag{2.4}$$

Since the labeling percentage and concentration of phenylalanine increased linearly over 6 h, $\frac{dC_{Phe}}{dt}$ and $\frac{df_{Phe}}{dt}$ were computed from the slopes of the time-series data. Then every 6-min estimates of C_{Phe} and f_{Phe} were derived based on $\frac{dC_{Phe}}{dt}$ and $\frac{df_{Phe}}{dt}$. f_{Tyr} was the average labeling percentage of tyrosine along the experiment. To determine the control v_c, emitted volatiles from day 2 control flowers fed with 10 mM tyrosine were collected for 2, 4, and 6 h starting at 18:00 h. No statistical differences were found in the scent profiles of flowers fed with 10 mM tyrosine compared to control. Since the volatile amount was found to increase linearly (Supplementary Fig. 4b in [79], v_c was assumed to be constant during the experimental period, and was derived from the slope of the time-series data. The v_c for *PhpCAT*-RNAi lines was subsequently determined by multiplying the control v_c by the average fractional decrease of total emission observed in lines 9 and 17 (Fig. 5a in [79]. With these obtained values, v_1 and v_2 were calculated along the experimental period every 6 min using equations 2.3 and

2.4. Variances in the estimated slopes were derived with a standard linear regression procedure as described in [156], while setting the intercepts as constants was based on experimental measurements. Since flux values are a function of the estimated trend slopes and other experimental measurements, flux variances can then be derived by considering the propagation of errors based on the following equation:

$$\sigma_y^2 = \sum_{i=1}^{n} \sigma_{x_i}^2 (\frac{\partial y}{\partial x_i})^2 \tag{2.5}$$

where $y = f(x_1, \ldots, x_n)$.

2.2.2 Metabolic flux modeling of aromatic amino acid biosynthesis

Aromatic amino acid biosynthesis was simplified as three parallel lumped reactions, each of which converts the common precursor glucose into Phe, Tyr or Trp. The equations for the three reactions focused on carbon mass balance are as follows:

$$1.67\ Glucose \xrightarrow{v_1} Phe + CO_2 \tag{2.6}$$

$$1.67\ Glucose \xrightarrow{v_2} Tyr + CO_2 \tag{2.7}$$

$$2\ Glucose \xrightarrow{v_3} Trp + CO_2 \tag{2.8}$$

Aromatic amino acids were assumed to be the major sinks for the shikimate pathway, and therefore its flux during the feeding studies was estimated from the sum of the fluxes through those three reactions. All fluxes were assumed to be constant within a 4-hour period of feeding time, and metabolite concentration dynamics in the pathway can be captured by linear functions, which matched with corresponding time course measurements. An empirical function was applied to simulate glucose labeling percentage dynamics and v_1 was constrained by known Phe-derived volatile emission rates. Inclusion of a metabolically inactive pool parameter for Phe improved fit with

experimental observations. In total, 11 unknown parameters were estimated for each case. Detailed involvements of these parameters into the model are listed below:

$$C_{Phe_{active,t}} = \theta_1 + \frac{\theta_2 - \theta_1}{4}t \tag{2.9}$$

$$C_{Tyr_t} = \theta_3 + \frac{\theta_4 - \theta_3}{4}t \tag{2.10}$$

$$C_{Trp_t} = \theta_5 + \frac{\theta_6 - \theta_5}{4}t \tag{2.11}$$

$$f_{Glucose_t} = \theta_7(1 - e^{-\theta_8 t}) \tag{2.12}$$

$$v_1 = v_{emission} + \frac{\theta_2 - \theta_1}{4} \tag{2.13}$$

$$v_2 = \theta_9 \tag{2.14}$$

$$v_3 = \theta_{10} \tag{2.15}$$

$$C_{Phe_{inactive}} = \theta_{11} \tag{2.16}$$

For each aromatic amino acid, labeling percentage dynamics were numerically integrated with the following balance equations [79] by ode15 solver in Matlab R2013a:

$$\frac{df_{Phe_{active,t}}}{dt} = \frac{(f_{Glucose_t} - f_{Phe_{active,t}})v_1}{C_{Phe_{active,t}}} \tag{2.17}$$

$$\frac{f_{Tyr_t}}{dt} = \frac{(f_{Glucose_t} - f_{Tyr_t})v_2}{C_{Tyr_t}} \tag{2.18}$$

$$\frac{f_{Trp_t}}{dt} = \frac{(f_{Glucose_t} - f_{Trp_t})v_3}{C_{Trp_t}} \tag{2.19}$$

To enable direct comparison of the models outputs with experimental measurements, the metabolically inactive Phe pool was integrated into the final outputs as follows:

$$C_{Phe_t} = C_{Phe_{active,t}} + C_{Phe_{inactive,t}} \tag{2.20}$$

$$f_{Phe_t} = f_{Phe_{active,t}} \frac{C_{Phe_{active,t}}}{C_{Phe_t}} \tag{2.21}$$

The objective function was defined as the differences between model-predicted profiles and experimentally measured profiles, weighed by the measurement variances. Parameters were estimated by minimizing the objective function through *lsqnonlin*

function in Matlab R2013a with multi-run approach. The mathematical representation of the optimization process is shown below:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum \frac{(Profile_{simulated} - Profile_{measured})^2}{s^2_{Profile_{measured}}} \tag{2.22}$$

Parameter uncertainty analysis was also performed as described in [156]. Briefly, 5,000 synthetic datasets were generated based on average and variance for each measurement by assuming a Gaussian distribution. For each dataset, the same optimization process was performed to obtain parameter values, and parameter uncertainty as well as model-predicted profile variance was estimated based on the variance of 5,000 parameter sets.

2.3 Results

2.3.1 PhpCAT controls phenylalanine biosynthetic flux [79]

Since downregulation of *PhpCAT* led to a decrease in the levels of phenylalanine, tyrosine, and their shared precursors prephenate and arogenate (Fig. 5a in [79], we hypothesized that phenylalanine and tyrosine accumulate inside plastids of *PhpCAT*-RNAi lines and feedback inhibit the arogenate pathway (Fig. 2.1a). Recently we showed that plants also contain an alternative pathway that proceeds via phenylpyruvate to produce phenylalanine in the cytosol, and flux through this route increases when the plastidial biosynthetic pathway (via arogenate) is impaired [7]. In the alternative pathway, a cytosolic PPY-AT preferentially converts phenylpyruvate to phenylalanine using tyrosine as an amino donor [7] (Fig. 2.1a). Interestingly, tyrosine cannot serve as an amino donor for PPA-AT in the plastidial arogenate pathway [157]. Taking advantage of this characteristic to distinguish between the two pathways, we employed metabolic flux analysis with stable isotopic labeling using ^{15}N-tyrosine to determine the effect of reduced plastidial phenylalanine export on carbon flux through the parallel phenylalanine biosynthetic pathways.

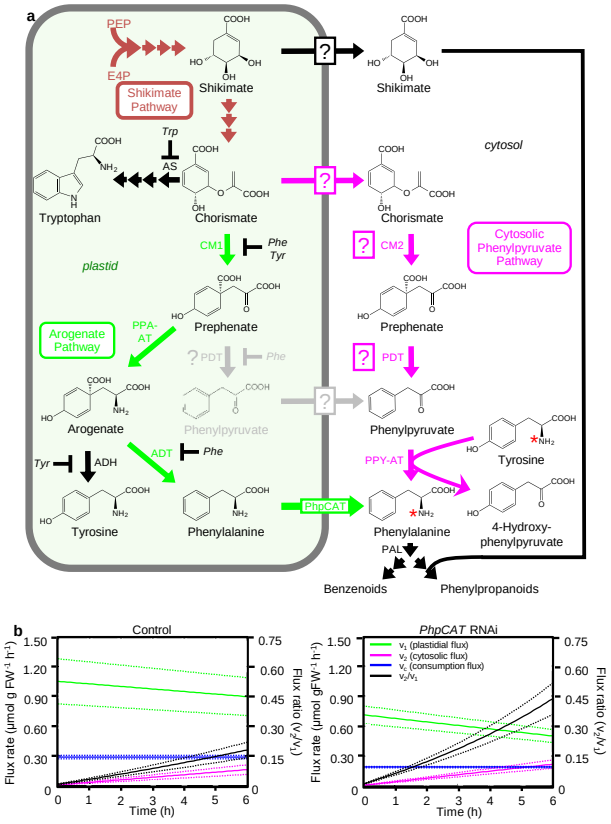

Fig. 2.1. Metabolic modeling of the phenylalanine biosynthetic network in control and *PhpCAT*-RNAi lines. (a) Scheme depicting cytosolic formation of phenylalanine and potential feedback inhibition mechanisms involved in plastidial phenylalanine biosynthesis [1]. White boxes with questions marks indicate unknown transporters/transport steps. AS, anthranilate synthase; E4P, erythrose 4-phosphate; PPA-AT, prephenate aminotransferase; PEP, phosphoenolpyruvate. (b) Flux models representing the phenylalanine biosynthetic network in 2-day-old petunia flowers from control and *PhpCAT*-RNAi lines. See Material and Methods for detailed modeling approach. v_1 flux through the plastidial arogenate pathway (green lines), v_2 flux through the cytosolic phenylpyruvate pathway (pink lines), v_c flux depicting the consumption of phenylalanine (blue lines). The v_2/v_1 ratio is shown by black lines. Solid lines are estimated values and dotted lines are standard deviation for each flux value. n = 3 for control and n = 6 for *PhpCAT*-RNAi lines.

Excised 2 day-old petunia flowers from control and *PhpCAT*-RNAi lines 9 and 17 were fed with 10 mM ^{15}N-tyrosine starting at 18:00 h, harvested after 2, 4, and 6 h, and analyzed by liquid chromatography-mass spectrometry (LC-MS) to determine phenylalanine and tyrosine pool sizes and isotope abundances. Similar to what was observed previously [7], the labeling percentage and concentration of phenylalanine increased linearly over 6 h, as did the level of emitted volatiles (Supplementary Fig. 4 in [79]. To assess the control PhpCAT exerts on metabolic fluxes through the phenylalanine biosynthetic network, a metabolic flux model was developed (See Materials and Methods). The simulation revealed that v_1, the rate of synthesis through the plastidial arogenate pathway, was 32 and 44% lower ($p < 0.05$, Students t-test, n$\geq$3, Bonferroni corrected) at t_0 and t_6, respectively, in *PhpCAT*-RNAi lines compared to control (Fig. 2.1b and Table 2.1). This finding is consistent with our hypothesis that reduced export of phenylalanine and tyrosine from plastids in *PhpCAT*-RNAi lines leads to feedback inhibition of the arogenate pathway. At the same time, the flux analysis showed that in control and the *PhpCAT*-RNAi lines, v_2, the flux through the cytosolic phenylpyruvate pathway, was minor at t_0, but significantly increased over the 6-h period (Fig. 2.1b and Table 2.1). In addition, increase in the rate of v_2 in *PhpCAT* knockdowns was more rapid than in control ($p < 0.01$, Students t-test, n$\geq$3, Bonferroni corrected) (Fig. 2.1b and Table 2.1), suggesting that more carbon flux is directed through the cytosolic pathway. Taking this into account with the decrease in flux through the arogenate pathway, the relative contribution of cytosolic phenylalanine production is considerably higher in *PhpCAT*-RNAi lines compared to control (a v_2/v_1 flux ratio of 0.44 versus 0.18, respectively, at t_6) (Fig. 2.1b and Table 2.1).

2.3.2 PhCM2 influences flux through both phennylalanine pathways [81]

To determine the effect of *PhCM2*-RNAi downregulation on the carbon flux through the phenylalanine biosynthetic pathways, 2-day-old control and *PhCM2* RNAi line

Table 2.1.
Plastidial arogenate (v_1) and cytosolic phenylpyruvate (v_2) pathway fluxes, and relative changes in flux, in flowers from control and *PhpCAT*-RNAi petunia lines at t_{0h} and t_{6h} (Shown as average±standard deviation).

	Plastidial synthesis rate v_1 (μmol g FW^{-1} h^{-1})			Cytosolic synthesis rate v_2 (μmol g FW^{-1} h^{-1})		
	t_{0h}	t_{6h}	Relative Change at t_{6h}	t_{0h}	t_{6h}	Relative Change at t_{6h}
Control	1.051±0.227	0.896±0.191	-14.7%	0.0067±0.0017	0.161±0.048	+2303%
PhpCAT-RNAi	0.711±0.089	0.500±0.065	-29.7%	0.0068±0.0010	0.218±0.041	+3106%
Relative change in *PhpCAT*-RNAi	-32.4%	-44.2%		+1.5%	+35.4%	

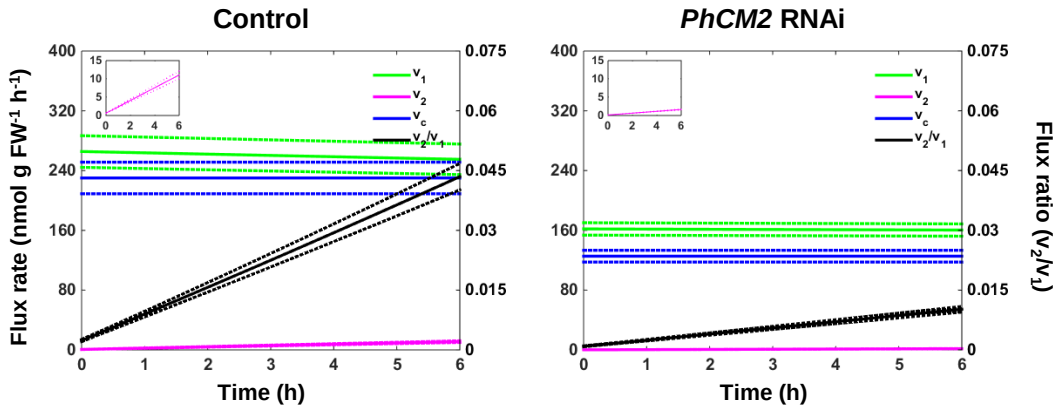

Fig. 2.2. Metabolic modeling of phenylalanine biosynthetic pathways in control and *PhCM2*-RNAi petunia flowers. Flux models representing the phenylalanine biosynthetic network in 2-day-old control and *PhCM2*-RNAi petunia flowers. See Material and Methods for detailed modeling approach. v_1 flux through the plastidial arogenate pathway (green lines), v_2 flux through the cytosolic phenylpyruvate pathway (pink lines, also in enlarged in inserts), v_c flux depicting the consumption of phenylalanine (blue lines). The v_2/v_1 ratio is shown by black lines. Solid lines are estimated values and dotted lines are standard deviation for each flux value (n = 3 biological replicates).

17 flowers were fed with 10mM ^{15}N-tyrosine starting at 18:00 h. After 2, 4 and 6 h of feeding, phenylalanine and tyrosine pool sizes and their isotopic abundances

Table 2.2.
Changes in flux in flowers from control and *PhCM2*-RNAi petunia lines at t_{0h} and t_{6h} (Shown as average±standard deviation).

	Plastidial synthesis rate v_1 (nmol g FW^{-1} h^{-1})			Cytosolic synthesis rate v_2 (nmol g FW^{-1} h^{-1})		
	t_{0h}	t_{6h}	Relative Change at t_{6h}	t_{0h}	t_{6h}	Relative Change at t_{6h}
Control	266±21.2	255±20.5	-3.95%	0.627±0.0560	11.1±1.08	+1660%
PhCM2-RNAi	162±8.28	160±8.22	-1.17%	0.147±0.0168	1.64±0.131	+993%
Relative change in *PhCM2*-RNAi	-39.0%	-37.3%		-76.2%	-85.2%	

were analyzed by LC-MS and used in our metabolic flux model of the phenylalanine biosynthetic network (See Materials and Methods for more details). As previously reported [7, 79], the labeling percentage of phenylalanine increased linearly over a 6 h period (Supplementary Fig. 6 in [81]. The model revealed that v_2, the flux through the cytosolic phenylpyruvate pathway, was 76 and 85% lower ($p < 0.0005$, two tailed Students t-test, n=3, Bonferroni corrected) at t_0 and t_6, respectively, in *PhCM2* RNAi plants relative to control (Fig. 2.2 and Table 2.2), further supporting that PhCM2 contributes to phenylalanine biosynthesis via the cytosolic phenylpyruvate pathway. At the same time, the flux analysis showed that in *PhCM2*-RNAi line, v_1, the flux through the plastidial arogenate pathway, was also decreased by 39 and 37% ($p < 0.005$, two tailed Students t-test, n=3, Bonferroni corrected) at t_0, and t_6, respectively, relative to control (Fig. 2.2 and Table 2.2). The decrease in plastidial aromatic amino acid biosynthesis is consistent with observed decrease in internal pools of arogenate and tryptophan (Fig. 2c in [81].

2.3.3 Reduction in PAL activity decreases carbon flux through the shikimate pathway and the accumulation of Phe that is metabolically inactive [80]

To determine changes in flux through the shikimate pathway, wild type and line 11 petunia flowers were fed with [U-$^{13}C_6$]-glucose, and pool sizes and isotopic abundances

Table 2.3.
Model-predicted metabolic fluxes within the aromatic amino acid biosynthetic network.

	Biosynthetic fluxes (nmol g FW^{-1} h^{-1})				
	Phe	Tyr	Trp	Shikimate	Inactive Phe (nmol g FW^{-1})
Wild type	456.6±91.4	0.14±0.02	0.04±0.01	456.8±91.4	188.5±26.8
PAL-RNAi	224.2±37.2	0.25±0.05	0.07±0.03	224.5±37.2	830.9±71.7
Relative change	-50.9%	+78.6%	+75.0%	-50.9%	+340.8%

All parameter means and variances were obtained based on the values estimated from 5000 synthetic datasets fit to experimentally determined label incorporation from exogenously supplied [U-$^{13}C_6$]-glucose over 4 h, as described in Materials and Methods.

of glucose, shikimate, and aromatic amino acids were analyzed at different time points over a 4-h period. The glucose labeling pattern was nearly identical in both control and transgenic petals (Fig. 2.3). In both genotypes, glucose labeling always exceeded labeling of shikimate, consistent with a simple precursor-product relationship between sucrose and shikimate (Figure S4 in [80]. Over this time course, the shikimate pool was significantly reduced in transgenic petals relative to the control, while there was little difference in its isotopic labeling (Figure S4 in [80]. We have constructed a dynamic model of the aromatic amino acid biosynthetic network to reproduce the observations from ^{13}C-labeled glucose feeding for both wild type and *PAL*-RNAi flowers (See Materials and Methods for more details). Shikimate pathway flux estimated by the model was found to be 50.9% less in *PAL*-RNAi than wild type (Table 2.3). There is a corresponding decrease of 50.9% in flux towards Phe, while the fluxes toward Tyr and Trp are increased by 78.6 and 75.0%, respectively.

The observed labeling patterns further suggested that within the cell there is an inaccessible largely unlabeled Phe pool, which based on the constructed dynamic flux model of the aromatic amino acid biosynthetic network was increased by 341% in the *PAL*-RNAi line 11 relative to wild type (Table 2.3).

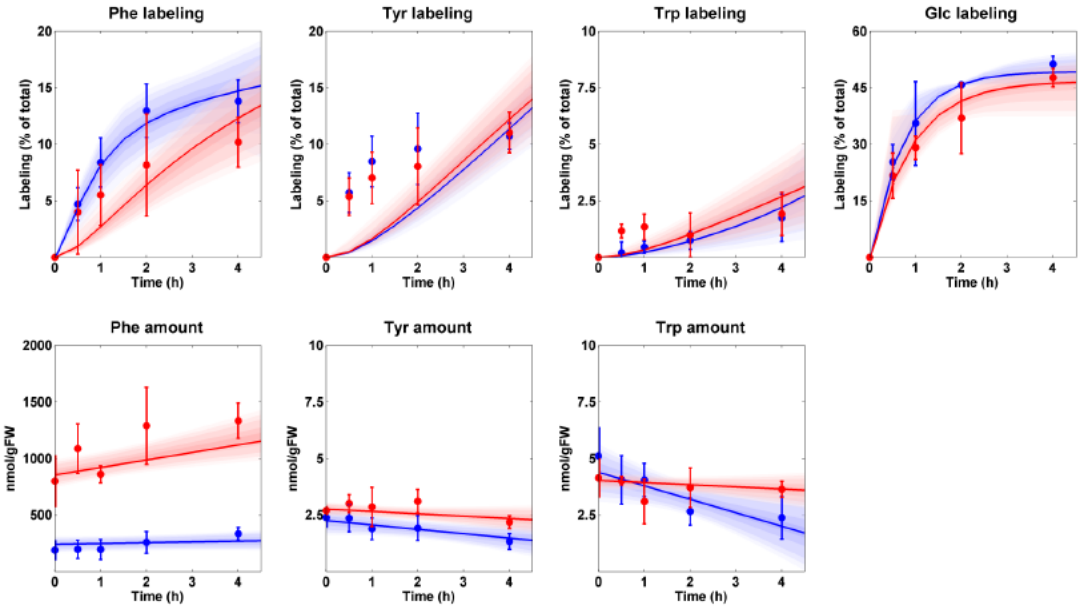

Fig. 2.3. Dynamic model simulation and experimentally obtained pool sizes and labeling patterns for aromatic amino acids and glucose in wild-type and *PAL*-RNAi petunia flowers. Isotopic abundances and pool sizes were analyzed over a 4-h time period (starting at 18:00 h) of [U-$^{13}C_6$]-glucose feeding to flowers of control (blue lines and symbols) and transgenic *PAL*-RNAi line 11 (red lines and symbols) plants. Lines represent simulation results, with the shaded area reflecting 95% confidence area for models outputs. Data points are the average of three biological replicates, error bars represent standard deviation.

2.4 Discussion

The purpose of the flux measurements presented here is to independently test the key genetic findings in the studies by [79, 80] and [81]. Due to lack of complete kinetic and metabolic information about the pathways, it is extremely difficult to construct an accurate mechanistic model to describe all the experimental results. Through carefully designed isotopic labeling studies, we were able to quantify metabolic fluxes

through each Phe biosynthetic branch using novel modeling strategies. As a result, we currently have a better understanding of Phe metabolism in plants.

[79] identified PhpCAT as a homologue of the *Escherichia coli* pheP [158]. PhpCAT is localized to plastids, and developmentally and rhythmically regulated just as other benzenoid biosynthetic genes with highest expression at night. It shows the transport activity for all three aromatic amino acids, and when its expression was reduced, significant decreases were observed for Phe and Tyr, with only minor decreases for Trp. Consistently, Phe and Tyr levels were increased in *PhpCAT*-overexpression lines, while Trp only showed an increasing trend (Fig. 5 in [79]. Our metabolic flux analysis indicated that these metabolite changes were closely correlated with plastidial shikimate pathway flux, as it decreased 32 and 44% at 18:00 h and midnight, respectively, in *PhpCAT*-RNAi lines compared to control. This flux decrease was partially compensated by the increased cytosolic flux, which contributed up to 30% of total Phe biosynthesis in *PhpCAT*-RNAi lines compared to 15% in control. The perturbation of the PhpCAT activity might mainly affect downstream steps in shikimate pathway instead of upstream ones, as shikimate amounts were only minimally affected in *PhpCAT*-RNAi and *PhpCAT*-overexpression lines, yet prephenate amounts were significantly perturbed as well as Phe and Tyr. Therefore, a possible scenario could be that the reduced activity of PhpCAT led to accumulated aromatic amino acids in the plastid. Potentially, feedback inhibition from these aromatic amino acids towards the downstream shikimate pathway enzymes caused accumulation of some pathway intermediates which were exported into the cytosol for cytosolic Phe biosynthesis. While it is not clear which intermediate was accumulated in this case, its elevated export into the cytosol could trigger the cytosolic phenylpyruvate pathway to dissipate the accumulated metabolic pools inside the plastid. In this way, plants can compensate the feedback control of plastidial aromatic amino acids when the plastidial export is impaired.

[7] has demonstrated the existence of a cytosolic phenylpyruvate pathway for Phe biosynthesis, yet the source of phenylpyruvate was unclear. [81] has characterized

PhCM2, a cytosolic enzyme involved in phenylpyruvate pathway, thus expanding the current knowledge of the cytosolic Phe biosynthesis. As expected from its metabolic role, suppression of *PhCM2* has led to the decreased prephenate amounts, as well as volatile emission and Phe amounts. Intriguingly, arogenate, Tyr and Trp amounts were also significantly decreased. These observations suggest that when the PhCM2 is impeded in the cytosol, the plastidial arogenate pathway was also affected. Our metabolic flux analysis also revealed the decreased flux values for both Phe biosynthetic pathways in *PhCM2*-RNAi lines, which not only indicates the direct involvement of PhCM2 into the cytosolic Phe biosynthesis, but also a potential yet unknown mechanism connecting the cytosol and plastid. This suggests that in spite of the low contribution of the cytosolic route to Phe biosynthesis, it might play an important role in maintaining Phe homeostasis.

Since all previous reported mechanisms of feedback control by Phe are effective in plastids, it remained to be addressed whether and how cytosolic Phe is being regulated in plants. [80] have generated *PAL*-RNAi lines in Petunia with 18-fold higher levels of Phe than wild type. Our flux model showed a decrease of 51% in biosynthetic flux of Phe in *PAL*-RNAi lines compared to wild type, which might be caused by the reduced shikimate supply, as shikimate amounts were significantly decreased. How the reduced cytosolic Phe consumption activity was coordinated with decreased plastidial shikimate pathway is not fully understood. One possible explanation is the accumulation of cytosolic Phe in response to impaired PAL activity feedback inhibited PhpCAT activity. This exporter inefficiency led to higher amounts of aromatic amino acids in plastids, which further shut down the shikimate pathway flux through allosteric feedback inhibitions. This possibility was supported by the observed high accumulation of both Tyr and Trp, although neither of their consumption were affected directly in *PAL*-RNAi lines. Our model also suggested a large metabolically inaccessible pool of Phe, which is 341% higher in *PAL*-RNAi lines compared to wild type. This pool of Phe was shown to be subcellularly separated from active cellular metabolism, evidenced by ^{13}C labeled carbon from glucose was not incorporated into this inactive Phe pool

during a 4-h feeding time period. Further experimental investigations suggested that this inactive Phe pool was likely sequestered in the vacuole. A tonoplast cationic amino acid transporter previously found in [79] was further characterized, and its ability to transport Phe was verified in a *Saccharomyces cerevisiae* expression system. When its activity was suppressed in Petunia *PAL*-RNAi lines, a significant increase in phenylacetaldehyde emission was observed. This could be due to the higher availability of cytosolic Phe since its vacuolar sequestration mechanism was impeded.

In conclusion, Phe biosynthesis in plants is regulated by different mechanisms to keep Phe availability at a sustainable level for its downstream consumption. These mechanisms not only include post-translational regulation, but very likely transcriptional regulation as well. Moreover, an alternative cytosolic Phe biosynthetic route is active in parallel with the tightly controlled plastidial pathway. This cytosolic route might be free from feedback inhibition of Phe, so that it plays a compensating role to maintain Phe homeostasis when the plastidial Phe export is impaired. While plastidial shikimate pathway enzymes are spatially separated from cytosolic Phe, upstream shikimate pathway might still be able to indirectly sense the cytosolic Phe level, probably by inhibition of the plastidial transporter by cytosolic Phe. On the other hand, in an extreme situation where Phe is hyperaccumulated in cytosol, a vacuolar sequestration mechanism is activated to isolate extra Phe from the rest of cellular metabolism. If one attempts to simulate metabolic dynamics of phenylpropanoid metabolism in plants, these regulatory mechanisms are key elements to be included for an accurate kinetic model.

3. DYNAMIC MODELING OF SUBCELLULAR PHENYLPROPANOID METABOLISM IN ARABIDOPSIS LIGNIFYING CELLS

Lignin is a polymer that significantly inhibits saccharification of plant feedstocks. Adjusting the composition or reducing the total lignin content have both been demonstrated to result in an increase in sugar yield from biomass. However, because lignin is essential for plant growth, it cannot be manipulated with impunity. Thus, it is important to understand the control of carbon flux towards lignin biosynthesis such that optimal modifications to it can be made precisely. Phenylalanine (Phe) is the common precursor for all lignin subunits and it is commonly accepted that all biosynthetic steps, spanning multiple subcellular compartments, are known, yet an in vivo model of how flux towards lignin is controlled is lacking. To address this deficiency, we formulated and parameterized a kinetic model based on data from feeding *Arabidopsis thaliana* basal lignifying stems with ring labeled [$^{13}C_6$]-Phe. Several candidate models were compared by an information theoretic approach to select the one that best matched the experimental observations. Here we present a dynamic model of phenylpropanoid metabolism across several subcellular compartments that describes the allocation of carbon towards lignin biosynthesis in wild-type Arabidopsis stems. Flux control coefficients for the enzymes in the pathway starting from arogenate dehydratase through 4-coumarate: CoA ligase were calculated and show that the plastidial cationic amino-acid transporter has the highest impact on flux.

3.1 Introduction

Plant lignocellulosic biomass is one of the most promising resources for second-generation biofuel production because of its ready availability and reduced environ-

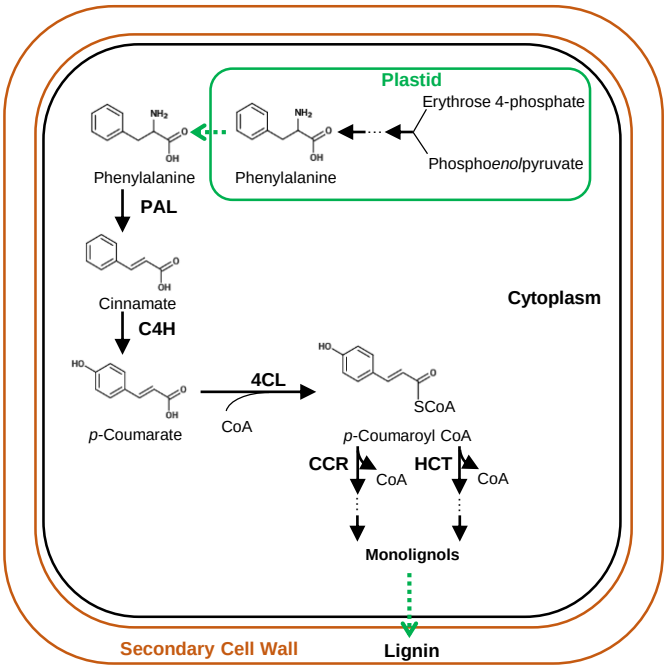

Fig. 3.1. Lignin biosynthesis in Arabidopsis. A simplified schematic of lignin biosynthesis is shown with metabolic steps explicitly simulated in the current study. Black arrows represent metabolic reactions with corresponding enzyme names. Green dashed arrows represent membrane transport steps between cellular compartments. Abbreviations: 4CL, 4-coumarate: CoA ligase; C4H, cinnamate 4-hyrdoxylase; CCR, cinnamoyl-CoA reductase; HCT, hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase; PAL, phenylalanine ammonia lyase.

mental impacts [159]. It is estimated that up to 442 billion liters of bioethanol could be produced each year from global lignocellulosic biomass [160], yet commercial application is largely inhibited by the complex production process that currently render it economically unfavorable [161]. One of the technical barriers lies in inefficient saccharification due to the presence of lignin in plants secondary cell wall. Lignin impedes the process by either physically restricting cellulose/hemi-cellulose accessibility [162], or adsorbing hydrolytic enzymes such as cellulase [16, 163]. An improvement in hydrolysis efficiency and a reduction in cost is thus expected by either reducing the total amount of lignin, or manipulating its composition [17, 20]. A better understanding of how

plants allocate carbon into lignin biosynthesis will be beneficial for these engineering efforts.

The general phenylpropanoid pathway connects primary metabolism with lignin biosynthesis in plants. All monolignols found in the lignin of dicotyledonous plants are derived from Phe, which is the key product of shikimate pathway and the precursor of phenylpropanoid compounds (Fig. 3.1). Multiple compartments are involved in phenylpropanoid metabolism, but genetic studies showed that Phe is predominantly synthesized inside the plastid, and translocated into cytosol with a cationic-amino acid transporter [4–6, 79]. Moreover, isotopic labeling experiments in petunia flowers revealed that 84.8% of Phe is produced inside the plastid during the period of active volatile production [79].

Plants possess several mechanisms to control flux through the phenylpropanoid network by regulating the synthesis of Phe and the activities of downstream phenylpropanoid enzymes, as well as by sequestering pathway intermediates. Arogenate dehydratase (ADT) in plastids, which catalyzes the conversion of arogenate to Phe, is known to be feedback inhibited by its product [11, 164]. Another mechanism includes post-translational regulation over PAL, the gateway enzyme for phenylpropanoid metabolism. The Kelch repeat F-box (KFB) protein family interacts with PAL thus mediating its ubiquitination and degradation [59–61]. In addition, cinnamate, the direct product of PAL, competitively inhibits PAL in several plant species [103, 165, 166]. Furthermore, plants utilize compartmental sequestration of pathway intermediates to control the flux through network. Indeed, nearly 90% of p-coumarate, the product of C4H was detected in the vacuole of soybean [167]. There is also recent evidence showing the importance of the vacuole in sequestering Phe under metabolic imbalance via a vacuolar Phe transporter [80, 168].

Kinetic modeling is a suitable approach to understand plant metabolism because it provides an integrative and quantitative framework [83, 97, 103, 115, 169]. Because of its mechanistic nature, once a valid kinetic model has been developed, metabolic control analysis can be applied to determine which step(s) within the network has the highest

control of flux [91, 103, 115]. In addition, manipulations can be explicitly simulated to predict possible outcomes of pathway perturbation [97, 113]. These predictions are helpful as they provide guidance for metabolic engineering efforts to reduce the number of genetic manipulations for experimental investigation [97].

Accurate model development involves selection of the best model(s) from multiple possible model candidates [170]. Akaikes Information Criterion (AIC) stems from the estimation of relative Kullback-Leibler distance of each model towards the ground truth, which is a fundamental information measure for any models. Thus, it provides an information theoretic way for model selection [171,172], which considers the balance between models performance and complexity. Using AIC and an *Escherichia coli* kinetic model, [173] successfully identified previously unknown allosteric interactions that are likely active *in vivo* in *E. coli*.

To date there are several mathematical models of lignin biosynthesis that have been developed by different groups [93, 94, 174]. [93] and [94] have applied generalized mass action kinetics to simulate each reaction within the pathway, and they observed that the model was able to fit the observations only when specific metabolic channels were considered. On the other hand, [174] simulated the pathway with Michaelis-Menten kinetics, and all kinetic parameters were measured by *in vitro* assays. Allosteric inhibitions towards enzymes from pathway intermediates were included in the model as well if the corresponding K_i values could be quantified. However, these models did not consider subcellular compartments and were not able to explore which of the regulatory mechanisms mentioned above contribute the most to control of carbon flux towards lignin biosynthesis. In this work, we developed a multicompartmental kinetic model connecting Phe biosynthesis with lignin production. By applying AIC, we identified a physiologically relevant kinetic model for general phenylpropanoid metabolism in Arabidopsis lignifying cells. The developed dynamic model quantitatively captures the critical regulatory properties of phenylpropanoid metabolism controlling carbon flux towards lignin biosynthesis.

3.2 Materials and Methods

3.2.1 Feeding study in wild type Arabidopsis stem

Arabidopsis thaliana (Columbia-0) wild type plants were grown in soil at 23 °C under 16/8-hour light/dark cycle. Feeding studies were conducted on primary stems of around 10 cm [175]. For each treatment, stems were excised and inserted into the medium with [$^{13}C_6$]-Phe (0.1, 0.3, 1 and 3 mM). [$^{13}C_6$]-Phe was purchased from Cambridge Isotope Laboratories, Inc. To avoid injury-induced artefacts due to cutting, only the segments 0.5-2 cm from the base of the stems were harvested, at time points of 0, 40, 90, 120, 180 and 240 min after insertion. Three replicates were collected and immediately frozen in liquid nitrogen for each time point, with 10 stem fractions pooled together as one replicate. Stem tissue was ground in liquid nitrogen with a mortar and pestle. An aliquot of each sample from time points at 0, 40, 90, 240 min were taken for enzyme assays. The rest of the sample was used for soluble phenylpropanoids analysis using LC/MS-MS [65] and lignin quantification by acetyl bromide lignin analysis and derivation followed by reductive cleavage (DFRC)/GC/FID/MS [176]. See sections below for detailed methods. Profiles of Phe and *p*-coumarate were directly applied for kinetic modeling. Sum of [$^{13}C_6$]-lignin monomers and all downstream intermediates of *p*-coumarate was used to represent the labeled product profile for each treatment.

3.2.2 LC-MS/MS analysis

The LC-MS/MS method developed by [65] was applied to quantify metabolites within the pathway. Briefly, tissue samples were extracted in 75% methanol (10 L per mg FW) and incubated for 2 hours at 65 °C. The samples were centrifuged for 10 min at 16,000 x g, and the supernatants were then dried in a speed-vac. Samples were re-dissolved in 60 L 50% methanol and analyzed using a Zorbax Eclipse C8 column (150 mm 4.6 mm, 5 μm, Agilent Technologies, Santa Clara, CA) as described in [65]. A QTrap 5500 triple quadruple mass spectrometer (AB Sciex, Redwood City,

CA) was used to detect soluble metabolites by multiple reaction monitoring in the negative ion mode. Quantification was achieved with calibration curve generated with standards of Phe, cinnamate, *p*-coumarate, caffeate, ferulate, sinapate, shikimate, *p*-coumaraldehyde, coniferaldehyde, sinapaldehyde, *p*-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol, *p*-coumaroyl shikimate, and caffeoyl shikimate. The same calibration was used to for both unlabeled and labeled isotopologues.

3.2.3 Lignin analysis

After soluble metabolites extraction, the cell wall debris was further washed with 10 mL 70% ethanol (v/v) five times and once with acetone. The cell wall residues were then dried at room temperature for lignin analysis. Total lignin was quantified using acetyl bromide lignin analysis following [177]. Briefly, about 2 mg cell wall residue was weighed and dissolved in 2.5 mL solution of acetyl bromide: glacial acetic acid (25:75) overnight at room temperature. Samples were mixed with 2.5 mL 2 M NaOH, 12 mL glacial acetic acid, and 0.5 mL 7.5 M hydroxylamine hydrochloride in a 50-mL volumetric flask. The final volume was brought to 50 mL with glacial acetic acid. Absorbance at 280 nm was measured for each sample and total lignin was calculated using the extinction coefficient of 23.29 g^{-1} L cm^{-1}. The rest of each cell wall residue sample was weighed and used for monomer composition analysis with DFRC/GC/FID/MS. Samples were dissolved overnight at room temperature in 2.5 mL of a solution of acetyl bromide in acetic acid (20:80) with 0.2 mg 4,4'-ethylidenebisphenol as internal standard. The products were dried then dissolved in 2 mL dioxane: acetic acid: water (50:40:10). 50 mg zinc dust was added into each sample and mixed well. After 20 min of reaction, about 200 L supernatants were loaded onto a LC-Si solid phase extraction column pre-conditioned with 95% ethanol and water and washed with 2.5 mL 25% ethanol. Samples were eluted with 2.5 mL 95% ethanol then dried before derivatization with 0.5 mL anhydrous acetic anhydride: pyridine (60:40) overnight. Acetylated products were then dried and dissolved in

200 μL dichloromethane. A 1 μL sample was analyzed on GC/FID for monomer composition and on GC/MS for labeling percentage of each monomer. To determine de novo lignin deposition rate, we harvested 0.5-2 cm stem sections from soil-grown plants each day for five days after bolting. Total lignin content in the sections was quantified with acetyl bromide method. Linear regression was applied to estimate the slope, which was used to determine the *de novo* lignin deposition rate. For labeled lignin quantification at each time point, we first measured the composition of each monomer released from lignin with DFRC/GC/FID. Then we estimated the amount of each monomer by multiplying total acetyl bromide lignin content. The labeled lignin content was calculated by multiplying the total amount of each monomer by its labeling percentage determined from DFRC/GC/MS analysis. The sum of all three types of labeled lignin was used to estimate the labeled lignin deposition rate.

3.2.4 Enzyme assays

Crude proteins were extracted from each sample with Tris-HCl buffer (pH 7.8) and desalted. PAL assays were performed in 100 mM Tris-HCl buffer (pH 7.5) with 5 mM Phe and 5 μL protein extract in a final volume of 50 μL, by incubating at 23 °C for 120 min. 4CL assays were performed in 100 mM Tris-HCl buffer (pH 7.5) with 5 mM $MgCl_2$, 5 mM ATP, 1 mM *p*-coumarate, 0.3 mM CoA and 2 μL protein extract in a final volume of 40 μL. Reactions were incubated at 23 °C for 20 min. Both assay products were analyzed with HPLC on a Shim-pack XR-ODS column (Shimadzu; column dimensions, 3 x 75 mm) using acetonitrile and 0.1% formic acid as mobile phases, and quantified using cinnamate and *p*-coumaroyl CoA as standards. Protein concentrations were determined by Bradford assay with bovine serum albumin as standard.

3.2.5 Kinetic modeling

The pathway is assumed to be at a metabolic steady state prior to external Phe feeding. The output flux from the pathway is set to be equal to the average lignin deposition rate (3.44 nmol g FW^{-1} min^{-1}) measured from 5-week-old Arabidopsis basal stems, as lignin is the major sink for the phenylpropanoid pathway in this type of tissue. The refined model also has two extra fluxes towards vacuole storage for both Phe and p-coumarate. The models stoichiometry is shown in Fig. 3.2 with abbreviations detailed in the caption. The mass balance equations for the refined model are:

$$\frac{d[^{12}Phe_{plastid}]}{dt} = {}^{12}v_{ADT} - {}^{12}v_{PCAT} \tag{3.1}$$

$$\frac{d[^{12}Phe_{cytosol}]}{dt} = {}^{12}v_{PCAT} - {}^{12}v_{PAL} - {}^{12}v_{VCAT} \tag{3.2}$$

$$\frac{d[^{13}Phe_{cytosol}]}{dt} = {}^{13}v_{PXT} - {}^{13}v_{PAL} - {}^{13}v_{VCAT} \tag{3.3}$$

$$\frac{d[^{12}Phe_{vacuole}]}{dt} = {}^{12}v_{VCAT} \tag{3.4}$$

$$\frac{d[^{13}Phe_{vacuole}]}{dt} = {}^{13}v_{VCAT} \tag{3.5}$$

$$\frac{d[^{12}pCA_{cytosol}]}{dt} = {}^{12}v_{PAL} - {}^{12}v_{4CL} - {}^{12}v_{PVT} \tag{3.6}$$

$$\frac{d[^{13}pCA_{cytosol}]}{dt} = {}^{13}v_{PAL} - {}^{13}v_{4CL} - {}^{13}v_{PVT} \tag{3.7}$$

$$\frac{d[^{12}pCA_{vacuole}]}{dt} = {}^{12}v_{PVT} \tag{3.8}$$

$$\frac{d[^{13}pCA_{vacuole}]}{dt} = {}^{13}v_{PVT} \tag{3.9}$$

$$\frac{d[^{13}Product]}{dt} = {}^{13}v_{4CL} \tag{3.10}$$

$$\frac{d[^{13}Phe_{xylem}]}{dt} = Volume_{xylem}[^{13}Phe_{medium}] \tag{3.11}$$

For the base model, mass balances of vacuolar metabolites and corresponding vacuolar transporter reactions are absent. For the models with cinnamate and/or *p*-coumaroyl CoA introduced, similar mass balance equations are introduced for both ^{12}C

and ^{13}C isotopes. The Phe concentration in the xylem is assumed to be equal with the Phe concentration in the feeding medium. Essentially this means the transpiration rate is assumed to be much faster than cellular uptake, so that the xylem is instantaneously filled with Phe upon feeding, and the concentration remains constant throughout the feeding period. The rest of reactions are modeled with Michaelis-Menten equations, with each isotopic molecule competing for the same enzymatic sites. Detailed reaction kinetics in the refined model are listed below:

$$^{12}v_{ADT} = \frac{v_{ADT}^{apparent}}{1 + \frac{[^{12}Phe_{plastid}]}{K_{i,Phe}^{ADT}}} \tag{3.12}$$

$$^{12}v_{PCAT} = V_{max,PCAT}\frac{[^{12}Phe_{plastid}]}{K_{m,Phe}^{PCAT}(1 + \frac{[^{12}Phe_{cytosol}]+[^{13}Phe_{cytosol}]}{K_{i,Phe}^{PCAT}}) + [^{12}Phe_{plastid}]} \tag{3.13}$$

$$^{13}v_{PXT} = V_{max,PXT}\frac{[^{13}Phe_{xylem}]}{K_{m,Phe}^{PXT} + [^{13}Phe_{xylem}]} \tag{3.14}$$

$$^{12}v_{PAL} = V_{max,PAL}\frac{[^{12}Phe_{cytosol}]}{K_{m,Phe}^{PAL} + [^{12}Phe_{cytosol}] + [^{13}Phe_{cytosol}]} \tag{3.15}$$

$$^{13}v_{PAL} = V_{max,PAL}\frac{[^{13}Phe_{cytosol}]}{K_{m,Phe}^{PAL} + [^{12}Phe_{cytosol}] + [^{13}Phe_{cytosol}]} \tag{3.16}$$

$$^{12}v_{VCAT} = V_{max,VCAT}\frac{[^{12}Phe_{cytosol}]}{K_{m,Phe}^{VCAT} + [^{12}Phe_{cytosol}] + [^{13}Phe_{cytosol}]} \tag{3.17}$$

$$^{13}v_{VCAT} = V_{max,VCAT}\frac{[^{13}Phe_{cytosol}]}{K_{m,Phe}^{VCAT} + [^{12}Phe_{cytosol}] + [^{13}Phe_{cytosol}]} \tag{3.18}$$

$$^{12}v_{4CL} = V_{max,4CL}\frac{[^{12}pCA_{cytosol}]}{K_{m,pCA}^{4CL} + [^{12}pCA_{cytosol}] + [^{13}pCA_{cytosol}]} \tag{3.19}$$

$$^{13}v_{4CL} = V_{max,4CL}\frac{[^{13}pCA_{cytosol}]}{K_{m,pCA}^{4CL} + [^{12}pCA_{cytosol}] + [^{13}pCA_{cytosol}]} \tag{3.20}$$

$$^{12}v_{PVT} = V_{max,PVT}\frac{[^{12}pCA_{cytosol}]}{K_{m,pCA}^{PVT} + [^{12}pCA_{cytosol}] + [^{13}pCA_{cytosol}]} \tag{3.21}$$

$$^{13}v_{PVT} = V_{max,PVT}\frac{[^{13}pCA_{cytosol}]}{K_{m,pCA}^{PVT} + [^{12}pCA_{cytosol}] + [^{13}pCA_{cytosol}]} \tag{3.22}$$

When the mass balances on either cinnamate or *p*-coumaroyl CoA are considered, the reaction kinetics for either C4H or CCR is introduced as a one-substrate Michaelis-Menten equation. With the presence of *p*-coumaroyl CoA, free CoA is modeled as

an independent substrate for 4CL kinetics, which leads to the new equation for 4CL reaction:

$$^{i}v_{4CL} = V_{max,4CL}\frac{[^{i}pCA_{cytosol}][CoA]}{K^{4CL}_{m,pCA}K^{4CL}_{m,CoA} + (K^{4CL}_{m,CoA} + [CoA])([^{12}pCA_{cytosol}] + [^{13}pCA_{cytosol}])}$$
$$i = 12 \; or \; 13 \tag{3.23}$$

In this case, net synthesis of CoA during the feeding study is assumed to be negligible, so that its mass balance is:

$$\frac{d[CoA]}{dt} = ^{12}v_{CCR} + ^{13}v_{CCR} - ^{12}v_{4CL} - ^{13}v_{4CL} \tag{3.24}$$

Each model with potential metabolite-enzyme interactions includes one possible regulatory interaction (activation, competitive or uncompetitive inhibition). The general form of such interactions in the corresponding enzymatic kinetics was modeled as:

$$v = V_{max}\frac{[activator]}{K_a + [activator]}\frac{[substrate]}{K_m(1 + \frac{[inhibitor]}{K_{ic}}) + [substrate](1 + \frac{[inhibitor]}{K_{iu}})} \tag{3.25}$$

3.2.6 Parameter identification

Not all model parameters are independent as steady state assumption is made at time 0 for the model. For the refined model, the number of independent parameters estimated is 15. Parameters were estimated by minimizing a pre-defined objective function score defined as the sum of square differences between models predictions and experimental observations weighted by measurement errors given training datasets:

$$\hat{sse} = \sum_{i=1}^{q}\sum_{t=1}^{T}\sum_{m=1}^{M}\frac{(Y^{predicted}_{m,t,i} - Y^{measured}_{m,t,i})^2}{s^2_{m,t,i}} \tag{3.26}$$

With M available measurements, each having T time points, and q different treatment datasets are used for model training.

The objective function was minimized first with *lsqnonlin* in MATLAB (R2013a, Mathworks, Inc), a gradient-based optimization algorithm with a multi-start approach

using random initial guesses. The best result was then used as the starting point for the following Markov Chain Monte Carlo (MCMC) sampling algorithm [108] to further explore the global minimum in the parameter space. For each MCMC batch, 100,000 samples were generated for an efficient exploration. The sampling was continued until the objective function score was not reduced with an additional MCMC batch.

For each parameter, the 95% confidence intervals were estimated by ranking 100,000 samples from highest to lowest and extracting values at $2{,}500^{th}$ and $97{,}500^{th}$.

3.2.7 Model comparison with information criterion

Models in the current study have different numbers of parameters. Therefore a direct comparison of the objective function fitting score is not appropriate since it is biased toward models with more parameters. Thus, AIC was applied to compensate for this issue [172]. The form of AIC score corrected for small sample/parameter ratio [172] is:

$$AIC = n \ln(\frac{s\hat{s}e}{n}) + \frac{2k(k+1)}{n-k-1} \tag{3.27}$$

where $s\hat{s}e$ is the sum of squared errors, which is the objective function score; n represents number of independent measurements, and k represents number of free parameters. A lower AIC score indicates an overall better model performance corrected with its complexity.

3.2.8 Metabolic control analysis

Metabolic control analysis provides a quantitative evaluation of the perturbation of enzyme amount towards pathway flux for each metabolic step [113]. The flux control coefficients (FCC) of an enzyme is mathematically defined as:

$$FCC_i = \frac{\partial J}{J} \Big/ \frac{\partial [E_i]}{[E_i]} \tag{3.28}$$

Here J represents flux through the pathway, and $[E_i]$ represents the target enzyme amount. Computationally, starting with a steady state model, V_{max} of the target

enzyme was changed by ±5% to get to new steady states, where new flux values were recorded which were used to compute FCCs.

3.3 Results

3.3.1 Metabolic profiling of wild-type Arabidopsis basal stem fed with ring labeled [$^{13}C_6$]-Phe

A feeding system in Arabidopsis stem was previously developed [175] using the basal portions of Arabidopsis stems in which the tissue segments were able to take up [$^{13}C_6$]-Phe from external medium, and incorporate it into downstream phenylpropanoids including lignin. Four feeding experiments were conducted using different [$^{13}C_6$]-Phe concentrations (0.1, 0.3, 1 and 3 mM) and metabolite profiling was performed at several points over a four-hour period (Fig. 3.3). The profiles describe the overall abundances of metabolites within the selected tissue, thus lack the subcellular details. On the other hand, significant labeling of extracted Phe was already achieved after 40 min of feeding with the degree of labeling increasing proportionally to the concentration of fed [$^{13}C_6$]-Phe (Fig. 3.3). After 40 min, labeled *p*-coumarate and lignin were readily detected in extracts from the stem tissue and the cell wall residue, respectively (Fig. 3.3 and Supplementary Fig. B.1), suggesting that the exogenous [$^{13}C_6$]-Phe is transported into stem tissue and efficiently incorporated into pathway end products.

To test if the activities of biosynthetic enzymes remain unchanged during the feeding process, the activities of PAL and 4CL in the stem tissue were determined after 0 min, 40 min, 90 min, and 240 min of feeding. As shown in Supplementary Fig. B.2, there were no changes in PAL and 4CL activities during the feeding process, independent of the exogenous Phe concentrations.

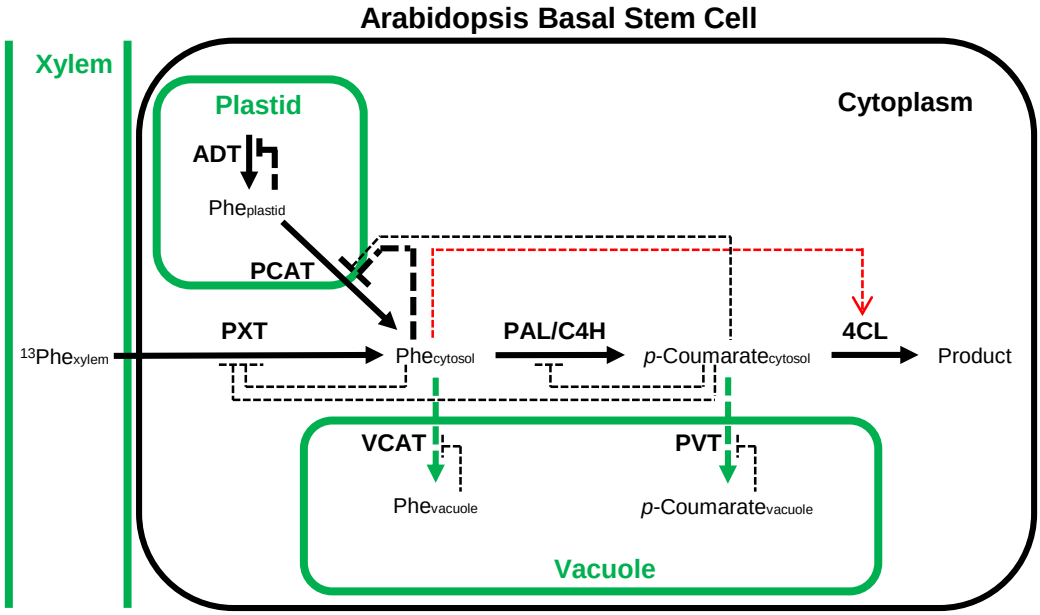

Fig. 3.2. Metabolite-enzyme interactions and vacuole storage mechanisms considered in the general phenylpropanoid pathway. The kinetic model in the current study is depicted with possible interactions between each metabolite and enzyme, and two possible vacuole storage mechanisms. Red dashed arrow represents feedforward activation, black dashed lines represent both competitive and uncompetitive feedback inhibition, green dashed arrows represent vacuole storage fluxes. Feedback competitive inhibition of plastidial phenylalanine towards ADT and cytosolic phenylalanine towards PCAT are bolded as they are included in the model a priori. Abbreviations: PCAT, plastidial cationic amino-acid transporter; PVT, putative vacuolar transporter; PXT, putative xylem transporter; VCAT, vacuolar cationic amino-acid transporter.

3.3.2 Base kinetic model construction and parameterization

A base kinetic model for the general phenylpropanoid pathway was constructed based on the known structure of phenylpropanoid metabolism in Arabidopsis [28]. A putative xylem transporter (PXT) was added to model Phe uptake from the xylem (Fig. 3.2). At the same time, the plastidial cationic amino-acid transporter (PCAT) was incorporated into the model to account for unlabeled Phe that is synthesized in

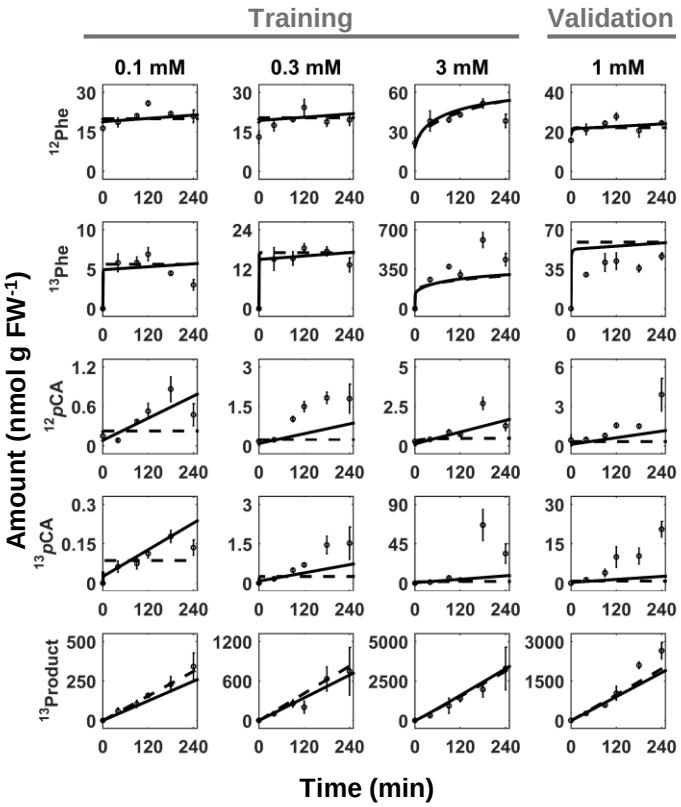

Fig. 3.3. Data fitting comparison between base and refined models. The model performance with experimental observations are shown for both base model (dashed) and refined model (solid) with vacuole storage for both phenylalanine and p-coumarate. Three treatments were used for both model training, with the 1 mM dataset left out for validation. Superscript 12 represents natural molecules with ^{12}C, while 13 represents molecules with ^{13}C labeled rings. ^{13}Product is the sum of all quantified phenylpropanoid molecules including lignin monomers labeled with ^{13}C. Details of experiments can be found in Materials and Methods. Measurements are the average ± standard deviation (n=3), while model predictions are shown as lines. Abbreviations: *p*CA, *p*-coumarate.

the plastid by ADT, and subsequently exported by transporter into the cytosol [79] where Phe is converted to cinnamate by PAL. As cinnamate was below the detection limit for all feeding experiments, it was not included in the base model. Therefore,

initially PAL and C4H enzymatic steps were lumped together into one single reaction. The product of this PAL/C4H reaction is *p*-coumarate, which is then consumed by 4CL *en route* to all monolignols in lignin biosynthesis. The base model treats this step as a one-substrate reaction due to lack of measurements of the other reactants, CoA and ATP. PAL and 4CL isoforms were not considered in the current model, as only total activity for these enzymes were measured. Therefore, all V_{max} and K_m values within the model are apparent ones. In addition to the enzymatic steps directly involved in phenylpropanoid metabolism, two feedback inhibition mechanisms were also included: (i) plastidial Phe competitive inhibition of ADT and (ii) cytosolic Phe inhibition of PCAT (Fig. 3.2). There were 11 parameters to estimate in the base model.

An efficient MCMC sampling algorithm was applied for parameter estimation [108]. To evaluate how well the proposed base kinetic model can explain the *in vivo* metabolite dynamics, three datasets with different levels of fed [$^{13}C_6$]-Phe (0.1, 0.3 and 3 mM) were first fit with the objective function to train the model, with the 1 mM dataset left out for model validation. Metabolite profiles predicted with estimated parameter values by MCMC for the base model were compared against experimental data (Fig. 3.3, dashed lines). Although the dynamics of Phe and downstream products were accurately captured by the model, it failed to predict *p*-coumarate trends in all cases (for both training and validation datasets), indicating that the base model was unable to fully capture the experimental observations. As the structure of the network included in the base model is experimentally supported in Arabidopsis [29, 31, 34, 35, 178], it is more likely that some regulatory mechanisms active *in vivo* remain unknown and thus had not been considered.

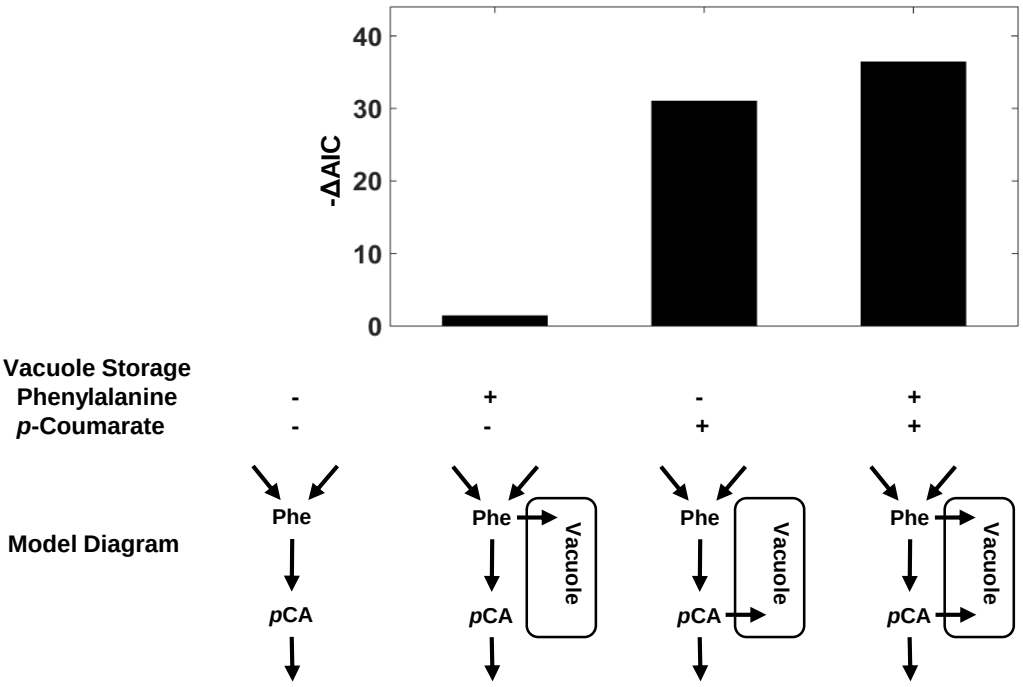

Fig. 3.4. Reduction in the AIC score relative to the base model without vacuole storage. Three additional models were generated and tested by considering possible combinations of vacuole storage mechanisms within the studied metabolism. Corrected AIC was applied to evaluate the performance of these models. The model with lowest AIC score is the one which fits the data best with the simplest structure.

3.3.3 Incorporation of vacuolar storage of Phe and *p*-coumarate significantly improve model performance

[80] identified a vacuole storage mechanism for Phe via a VCAT in petunia. Arabidopsis contains four homologues encoding vacuolar membrane amino acids transporters [179, 180]. Additionally, subcellular fractionation studies have detected the vacuolar pools of Phe [168] and *p*-coumarate [167] *in planta*. Therefore, we hypothesized that the vacuole storage mechanisms are also present in Arabidopsis lignifying cells. The putative vacuolar transporters could remove excessive Phe and *p*-coumarate from the cytosol into this separate compartment, to maintain lignin

biosynthetic homeostasis if the concentration of either compound changes substantially. To test this hypothesis, we added two vacuolar transporters into the base model. Because simulation had been performed over a short time period (4 hour), the vacuole transporters were modeled to function unidirectionally from cytosol to vacuole. Thus, each transporter can be modeled with two additional parameters (See Materials and Methods for detailed equations). Three competing models were formulated with vacuole storage mechanisms for Phe and/or *p*-coumarate (Fig. 3.4). To evaluate which, if any, of these competing models have better performance, the AIC corrected for low sample/parameter ratio was applied to all four datasets. The model with storage mechanisms for both Phe and *p*-coumarate had the largest reduction in AIC score relative to the base model, indicating that having those additional transporters significantly improved the fit (Fig. 3.4).

To evaluate whether the refined model has prediction capability outside of training data, the refined model with two storage mechanisms was trained using three datasets (0.1, 0.3 and 3 mM), while the 1 mM dataset was used for model validation, just as it was done for the base model. As expected, the refined model better matched the data trends in both the training and validation sets as compared to the base model (Fig. 3.3), especially on the predictions of *p*-coumarate profiles. The simulations of the refined model suggest that *p*-coumarate mainly accumulates inside the vacuole during feeding studies (Table 3.1), providing the evidence of an active sequestration mechanism when Phe was present at high concentrations.

Next, all four datasets were utilized for the final parameter identification. Estimated parameters were compared with literature values when available (Table 3.2). The model accurately predicted PAL K_m for Phe which within a 95% confidence interval was consistent with *in vitro* measurements (3.3 versus 4.3 nmol g FW^{-1}). However, the 4CL K_m for *p*-coumarate was largely underestimated (0.30 versus measured 2.4 nmol g FW^{-1}). Since 4CL step was modeled as a single-substrate reaction, it should be noted that the estimated K_m is a combined one including the influence of CoA and ATP. A significantly lower estimated value might indicate the either CoA and/or ATP

Table 3.1.
Subcellular metabolite profiles under different feeding conditions predicted by the refined model (Units: nmol g FW^{-1}, values are represented as average ± standard deviation estimated by MCMC).

Condition	t = 0 min	t = 240 min			
Metabolite (Location)	Non-fed	0.1 mM	0.3 mM	1 mM	3 mM
Phe (Xylem)	0.0±0.0	2.8±0.3	8.4±0.8	27.8±2.6	83.5±7.7
Phe (Plastid)	15.4±0.9	15.4±0.8	15.5±0.8	16.1±0.8	25.0±2.1
Phe (Cytosol)	1.7±0.6	2.3±0.8	4.0±1.3	12.6±3.9	220±21
Phe (Vacuole)	0.0±0.0	9.4±1.0	11.2±1.2	14.3±2.7	16.4±4.3
p-coumarate (Cytosol)	0.07±0.01	0.10±0.01	0.16±0.02	0.40±0.04	1.11±0.13
p-coumarate (Vacuole)	0.00±0.00	0.93±0.04	1.5±0.1	3.7±0.2	9.3±0.7

pool *in vivo* is not saturating 4CL, as occurs in *in vitro* assays. Consistently, 4CL maximal capacity predicted by the model was significantly lower than 4CL activity measurements from crude extracts (Table 3.2 and Supplementary Fig. B.2), indicating that 4CL was not functioning at its in vitro maximal velocity likely due to low CoA and/or ATP concentration.

3.3.4 No *in silico* evidence for existence of other metabolite-enzyme interactions

The proposed models (both base and refined ones) included two competitive inhibitions based on the previous knowledge of the pathway [11, 79, 164]. These models have the risk of being biased against unknown factors because *a priori* no other metabolite-enzyme regulatory interactions were considered. The wide range of perturbations undertaken in the feeding studies (from 0.1 up to 3 mM treatment) provides an excellent dataset to search for any significant regulatory interactions. Starting with the refined model, 14 possible metabolite-enzyme interactions were evaluated

Table 3.2.
Final parameter list estimated by the refined model.

Parameter identity	Reference value	Estimated value	95% Confidence intervals[a]	Unit
Total Phe amount at t = 0	16.7±3.9[b]	17.2	[16.3, 17.8]	nmol g FW^{-1}
Total *p*CA at t = 0	0.3±0.2[b]	0.07	[0.06, 0.09]	nmol g FW^{-1}
Cytosolic fraction of Phe at t = 0	-	0.06	[0.04, 0.18]	dimensionless
Xylem volume	-	31.7	[22.6, 32.6]	μL g FW^{-1}
$K_{m,Phe}^{PCAT}$	-	5.0	[0.4, 9.8]	nmol g FW^{-1}
$K_{m,Phe}^{PXT}$	-	56.3	[45.6, 76.9]	nmol g FW^{-1}
$K_{m,Phe}^{PAL}$	4.3[c]	3.3	[2.2, 12.0]	nmol g FW^{-1}
$K_{m,Phe}^{VCAT}$	-	1.2	[0.17, 4.3]	nmol g FW^{-1}
$K_{m,pCA}^{4CL}$	2.4[d]	0.30	[0.27, 0.48]	nmol g FW^{-1}
$K_{m,pCA}^{PVT}$	-	70.0	[20.7, 98.9]	nmol g FW^{-1}
$K_{i,Phe}^{ADT}$	-	0.10	[0.10, 0.73]	nmol g FW^{-1}
$K_{i,Phe}^{PCAT}$	-	34.5	[5.2, 94.8]	nmol g FW^{-1}
$V_{max,ADT}$	-	548	[75.1, 537]	nmol g FW^{-1} min^{-1}
$V_{max,PCAT}$	-	4.6	[3.6, 5.7]	nmol g FW^{-1} min^{-1}
$V_{max,PXT}$	-	20.4	[18.9, 28.6]	nmol g FW^{-1} min^{-1}
$V_{max,PAL}$	10.5±3.0[b]	14.7	[14.3, 17.5]	nmol g FW^{-1} min^{-1}
$V_{max,VCAT}$	-	0.07	[0.04, 0.11]	nmol g FW^{-1} min^{-1}
$V_{max,4CL}$	82.4±21.4[b]	19.0	[18.0, 23.4]	nmol g FW^{-1} min^{-1}
$V_{max,PVT}$	-	3.0	[0.74, 4.2]	nmol g FW^{-1} min^{-1}

[a] Estimated with 100,000 MCMC samples;

[b] Measured in current study;

[c] [34], cytosol volume setimated from [181];

[d] [29].

(Fig. 3.2, Supplementary Table B.1). These included all possible combinations between metabolite and enzymes interactions: competitive inhibition, uncompetitive inhibition and feedforward activation of enzymes by metabolites. The AIC score was applied to evaluate if a model with any of the proposed interactions had a better performance than the refined model. Since no reduction in AIC score was observed for any model, the data do not support the *in vivo* existence of any tested metabolite-enzyme interactions (Supplementary Table B.1).

3.3.5 Expanding the model with additional metabolites did not improve models performance

Cinnamate was not initially included as an independent metabolite in the model because its amount was below the detection limit with LC-MS/MS [65]. Yet it is possible that its presence could have significant impact on the models behavior,

considering that cinnamate inhibits PAL activity in plants including Arabidopsis [165]. Additionally, CoA, a substrate for 4CL-catalyze reaction was also not included in the model because of lack of information about its subcellular levels. The product of 4CL is *p*-coumaroyl CoA, which is a precursor for either CCR *en route* to the synthesis of H lignin, or HCT towards production of G and S lignin (Fig. 3.1). Both enzymes release the CoA moiety back to the free CoA pool. Since underestimation of the 4CL K_m for *p*-coumarate suggested the possible limitation of CoA *in vivo*, dynamics of CoA pools might play a role in regulating flux through 4CL step. To test the effect of these metabolites, three additional models including the mass balances of cinnamate and/or *p*-coumaroyl CoA (the latter also includes CoA balance) were constructed and fit to the data from all four feeding experiments for comparison against the refined model. As shown in Supplementary Fig. B.3, including a cinnamate mass balance into the model did not improve the statistical fit. This suggests that C4H is not a limiting step even during feeding experiments with high concentrations of Phe. Incorporation of the *p*-coumaroyl CoA mass balance also did not improve models performance, suggesting that within those feeding treatments CoA balance did not have significant impact on 4CL flux (Supplementary Fig. B.3).

3.3.6 Refined model unravels subcellular pathway dynamics during feeding studies

All attempts described above to improve the model revealed that the kinetic model with vacuole storage mechanism for Phe and *p*-coumarate (refined model) best represents *in vivo* situations. This model was then used to quantify two Phe input fluxes via PXT and ADT and the output flux through 4CL under different feeding conditions (Fig. 3.5). The [$^{13}C_6$]-Phe uptake rate from the xylem (via PXT) into the lignifying cells in the basal stems continuously increased with higher fed Phe concentration up to 12 nmol g FW^{-1} min^{-1}, thus contributing from 24% up to 89% of flux that reaches 4CL at 0.1 mM and 3 mM of fed Phe, respectively (Fig. 3.5). On

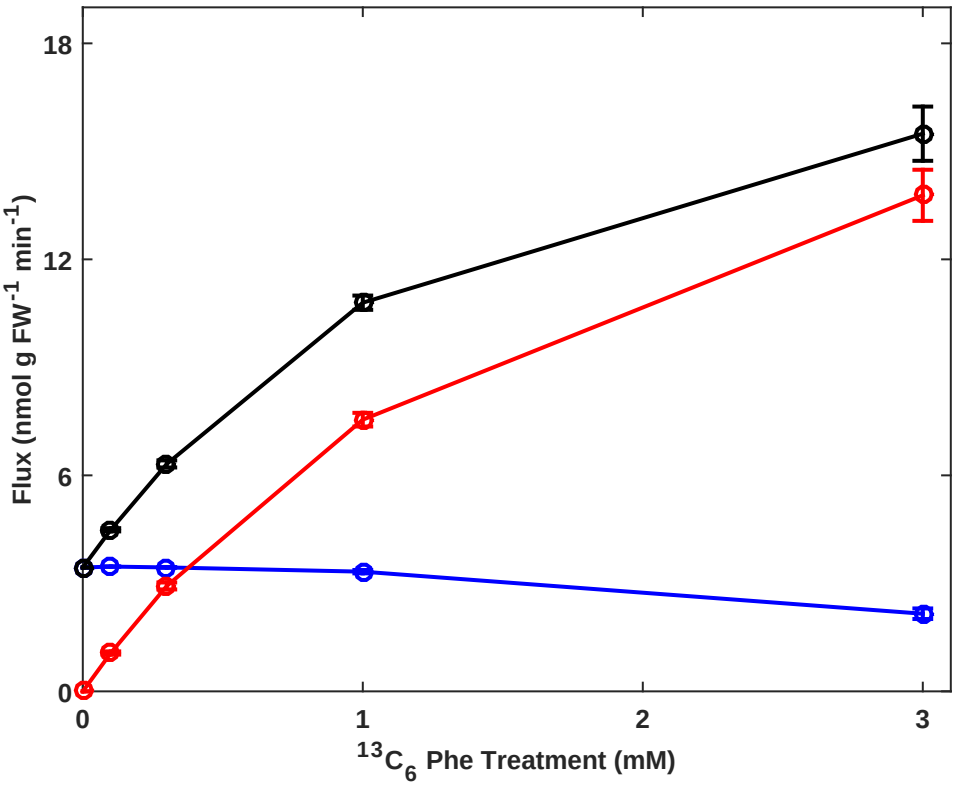

Fig. 3.5. Pathway flux distribution at 240 min for different feeding treatments. Fluxes at 240 min for different feeding treatments were predicted by the refined model. Shown were the average ± standard deviation values for fluxes from 100,000 MCMC simulation samples. Black line represents total carbon flux through the 4CL enzyme (sum of ^{12}C and ^{13}C fluxes), red line is the flux through the PXT, and the blue line is the reaction rate of ADT.

the other hand, accumulation of cytosolic Phe at high fed Phe concentrations resulted in a decrease in the plastidial Phe synthesis rate (Fig. 3.5 and Table 3.1).

Analysis of model-predicted subcellular distribution of Phe and *p*-coumarate revealed that at low Phe fed concentrations (0.1 mM and 0.3 mM) the enzymes in the cytosol are capable of efficiently utilizing increasing amounts of substrates, as both Phe and *p*-coumarate were around or below their respective K_m for PAL and 4CL, respectively (Table 3.1 and 3.2). Vacuole storage mechanisms played important roles for pathway homeostasis since they sequester a significant amount of Phe and

p-coumarate (29% and 90% out of total for each, at 0.3 mM fed Phe) (Table 3.1). However, at higher fed Phe concentrations (1 mM and 3 mM) the model predicted that the vacuolar transporters became saturated, resulting in dramatic expansion in the cytosolic Phe and *p*-coumarate pools. After 240 min, cytosolic Phe and *p*-coumarate pools were 16.4-fold and 1.8-fold higher respectively at 3 mM fed Phe than at 1 mM fed Phe, suggesting that PAL and 4CL were no longer able to efficiently convert these substrates into downstream products. The hyper-accumulated cytosolic Phe can feedback inhibit PCAT leading to the build-up of plastidial Phe, which in turn inhibits ADT. Interestingly, while an accumulation of Phe was observed in the cytosol (17.5-fold), the plastid Phe level for the 3 mM fed Phe only increased by 1.6-fold compared to 1 mM fed Phe, mainly due to tight feedback regulation of ADT (Table 3.1).

3.3.7 Metabolic control analysis determines the relative enzymatic control of flux in the general phenylpropanoid pathway

The FCC determined by metabolic control analysis (MCA) provide a quantitative view of the distribution of control that the individual enzymes impose on each flux within the metabolic network. Under non-fed Phe conditions, the highest control coefficients were found upstream of PAL (Table 3.3). Specifically, PCAT had the highest control coefficient over the flux, which is consistent with the role of petunia plastidial cationic amino-acid transporter (PhpCAT) in phenylpropanoid metabolism in flowers [79]. Indeed, a 75-80% downregulation of *PhpCAT* transcript levels led to 20-42% reduction in the phenylpropanoid pathway flux. MCA also indicates that downstream enzymes do not have significant controls over the pathway flux in Arabidopsis stems under non-fed conditions. When PAL and 4CL activities were measured in crude extracts of basal stem tissue used in the feeding studies, both values were significantly higher than lignin deposition rate (on average over all feeding conditions 10.5 nmol g FW^{-1} min^{-1} for PAL, 82.4 nmol g FW^{-1} min^{-1} for 4CL, and

Table 3.3.
Flux control coefficients of pathway enzymes.

Enzyme	Flux control coeffcient	95% Confidence intervals[a]
ADT	0.19	[0.03, 0.28]
PCAT	0.80	[0.71, 0.97]
PAL	0.01	[0.01, 0.02]
4CL	0.00	[0.00, 0.00]
VCAT	-0.01	[-0.01, -0.01]
PVT	0.00	[0.00, 0.00]

[a]Estimated with 100,000 MCMC samples.

3.4 nmol g FW^{-1} min^{-1} for lignin synthesis; Table 3.2 and Supplementary Fig. B.2), suggesting that neither enzyme was functioning at their V_{max} levels. Together with the very low estimated FCCs, these observations suggest that phenylpropanoid metabolism is mainly controlled by Phe biosynthetic reactions and its export from the plastid.

3.4 Discussion

This study developed multiple kinetic models of the phenylpropanoid metabolism in Arabidopsis lignifying cells by evaluating different regulatory mechanisms. As plant metabolism is highly compartmentalized [182], subcellular compartments were also integrated in the model allowing estimation of subcellular concentrations of metabolites. Cascade feedback inhibition of plastidial ADT by cytosolic Phe via PCAT was incorporated in the current kinetic model of Arabidopsis phenylpropanoid metabolism. The refined model with this inhibition was able to accurately capture Phe profiles in all feeding datasets. Recently, it has been suggested that the plant is able to sense the amount of cytosolic Phe and control carbon allocation to its biosynthesis [80]. Indeed, hyperaccumulation of Phe in *PAL*-RNAi petunia petals and Arabidopsis double *pal1pal2* and quadruple *pal1pal2pal3pal4* knockout mutants

reduced carbon flux in the shikimate pathway, which provides precursors for Phe biosynthesis [80]. In fact, without feedback control of plastidial Phe biosynthesis by cytosolic Phe, no model is able to explain the observed accumulation of unlabeled plastidial Phe upon feeding, especially at 3 mM fed Phe (Fig. 3.3) and reduction of flux through ADT, which has a low K_i value for Phe (0.10 nmol g FW^{-1}, Table 3.2). Indeed, a drastic change of cytosolic Phe concentration was predicted for 1 mM, and 3 mM fed Phe, while changes in plastidial Phe concentrations were relatively small (Table 3.1). The cytosolic Phe concentration (220 nmol g FW^{-1}, Table 3.1) was far above the predicted PCAT K_i value for Phe (34.5 nmol gFW^{-1}, Table 3.2), suggesting that during hyper-accumulation of cytosolic Phe, the plastidial Phe transporter will be feedback inhibited.

An information-theoretic approach was applied to select the model with the best performance, which suggested vacuole storage for both Phe and *p*-coumarate. Although initial pools of Phe and *p*-coumarate inside the vacuole before feeding are not identifiable through current approach, these storage mechanisms seem to function upon the feeding of exogenous Phe. It is thus hypothesized that active vacuolar transporters help maintain the flux homeostasis by translocating excess amounts of substrate/intermediates into a physically separate compartment, making them unavailable for enzymes. The role of the vacuole in sequestration of excess cytosolic Phe was recently demonstrated by [80]. Phe amount inside the vacuole was nearly doubled in *PAL*-RNAi petunia relative to wild-type control likely due the block of cytosolic Phe consumption.

Despite the low fluxes into the vacuole relative to the main flux towards lignin (Supplementary Fig. B.4), the optimized model predicts that substantial accumulation of Phe and *p*-coumarate occurs in the vacuole (Table 3.1). Indeed, the estimated maximal capacities for both vacuolar transporters are smaller (0.07 and 3.0 nmol g FW^{-1} min^{-1}) than the other V_{max} values within the pathway (Table 3.2). The low capacity of Phe transporter was insufficient to efficiently translocate cytosolic Phe under high Phe (1 and 3 mM) feeding, which is the main cause of hyperaccumulation

of cytosolic Phe. The most similar homologues of petunia PhCAT2 in Arabidopsis are *AtCAT2* and *AtCAT4* in stems, encoding vacuolar amino acids transporters [180]. Under normal growth conditions, they have relatively low expressions [180, 183]. However, under abiotic and biotic stresses which are known to result elevated free Phe pools in plants, the expression of *AtCAT2* and *AtCAT4* are induced [183, 184], suggesting these transporters mediate the transfer of excess Phe into vacuole.

The control coefficients of ADT, PCAT, PAL, C4H and 4CL were calculated and the results suggest PCAT has the most significant control on flux towards lignin under non-fed conditions. It has been shown previously that perturbing the expression of the plastid transporter leads to significant changes in the downstream efflux in petunia flowers [79]. The reason for this translocation step to be limiting is because of its relatively low predicted K_m value compared to the plastidial Phe amount (5.0 versus 15.4 nmol g FW^{-1}, Table 3.1 and 3.2). No *in vitro* enzyme assay has been done to characterize the K_m value for PCAT in Arabidopsis, although a homolog in *Escherichia coli* has a K_m value of 2 μM [158], corresponding to around 0.4 nmol g FW^{-1} based on the estimated plastid volume in plants [181]. Both calculations suggest that this transporter is easily saturated under normal condition, so that the translocation rate is largely determined by the protein amount instead of substrate level. For the other enzymatic steps, the model suggests that they are controlled by substrate levels instead of protein amounts, which explains why pathway flux is insensitive to small changes of enzyme levels in MCA. However, when a large perturbation towards the enzyme levels occurs, eventually the control will be shifted from substrate to enzyme.

Cinnamate is involved in the phenylpropanoid metabolism as a product of PAL and is the substrate for *p*-coumarate synthesis. During the feeding studies, especially at 3 mM treatment, both Phe and *p*-coumarate accumulate to a much higher amount than under the non-fed condition, suggesting that [$^{13}C_6$]-Phe has been effectively taken up by the plant, and the pathway capacity has reached a saturation point in converting those metabolites to lignin. Even under this condition, no cinnamate was detected in extracts. Inclusion of a cinnamate mass balance into the model did not improve

the models performance, while the model was able to fit the data well by simulating PAL and C4H as one single reaction. These observations might be explained by a very efficient (high k_{cat}/K_m) C4H enzyme within the pathway. As a membrane bound protein, it is very difficult to directly test this hypothesis by measuring its maximal capacity in the basal stem tissue. Metabolic channeling has been proposed between PAL and C4H [185, 186], which provides another possible explanation of undetectable cinnamate, since the intermediate could then be efficiently passed between enzymes without having any significant accumulation.

We also attempted including mass balances for CoA and *p*-coumaroyl CoA, but found the model fits unimproved. Although the current refined model still has some underfitting issues over *p*-coumarate profiles for the 3mM treatment, it was not resolved by modeling 4CL step as a bi-substrate reaction. Another factor that might have significant effect but was not considered is ATP, which is also required by the 4CL reaction. As ATP has multiple subcellular locations, a metabolomics technology with the ability to estimate its subcellular concentrations is required.

In summary, kinetic modeling combined with a statistical evaluation procedure successfully described metabolite distributions and fluxes within the phenylpropanoid network in Arabidopsis lignifying cells. The AIC model selection procedure supported by feeding treatment datasets proposed regulatory mechanisms active *in vivo*, which could be experimentally tested in future. In addition, MCA suggested limitations within the phenylpropanoid network, thus providing a useful guide for future metabolic engineering efforts. For example, knocking down phenylalanine export from the plastids is a first metabolic engineering target for reducing lignin biosynthesis.

4. COMBINING GENE REGULATORY NETWORK AND KINETIC MODELING OF LIGNIN BIOSYNTHESIS IN ARABIDOPSIS

4.1 Introduction

Lignin is a heterologous biopolymer which is deposited in secondary cell walls of terrestrial plants. It is one of the most abundant cell wall polymers and is crosslinked with two other major polymers, cellulose and hemicellulose to form a complex matrix [187, 188]. Studies have shown that lignin plays important roles in normal plant growth, by contributing to the lodging resistance [189], the water proofing of conductive elements in xylem [190,191], as well as against external hazards [192,193]. On the other hand, plant lignocellulosic biomass has drawn lots of attention for its great capacity and reduced environmental impacts as the source for second-generation biofuels [159]. Current industrial applications of lignocellulosic biomass are limited by the high costs during the pre-treatment of cell wall materials, and lignin is reported as one of the major inhibitory factors due to its chemical and structural characters [187, 194]. Therefore, one of the major strategies to improve biomass pretreatment is to reduce lignin content or manipulate its composition [17], and a systematic understanding of the lignin biosynthesis in plants is necessary for the success of such strategies.

Lignin is mainly composed of *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) subunits, which are derived from *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, respectively. In certain cases and in specific tissues, other types of subunits are observed [23, 32, 195, 196], which are derived from other aromatic aldehydes/alcohols that share the common precursor L-phenylalanine (Phe). A series of enzyme families are active in lignifying cells to convert Phe to downstream monolignols, with isoforms existing for phenylalanine ammonia lyase (PAL), 4-coumarate: CoA ligase (4CL) and

cinnamyl alcohol dehydrogenase (CAD) in Arabidopsis (Fig. 4.1). The presence of isoforms provides redundancy for pathway robustness, since larger decreases in lignin content were frequently observed when multiple isoforms were impaired simultaneously, compared with the cases when only one of isoforms was perturbed [23, 31, 35]. In addition, both 4CLs and CADs are capable of catalyzing multiple reactions within the pathway, due to similar molecular structures of their substrates (Fig. 4.1). Enzyme promiscuities are also present for cinnamoyl-CoA reductase (CCR, [36, 37], aldehyde dehydrogenase (ALDH, [38], hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT, [39], ferulate 5-hydroxylase (F5H, [40] and caffeic acid O-methyltransferase (COMT, [41] (Fig. 4.1). The presence of isoforms and substrate competitions complicates the metabolic network of lignin biosynthesis, which leads to numerous possible routes between Phe and downstream monolignols. These complexities make it difficult to predict the consequences of genetic perturbations for some metabolic steps within the pathway, since unexpected outcomes might be triggered by events such as distal interactions between competing substrates, activation of hidden routes upon pathway block, etc. While several successful manipulations have been made in lignin biosynthesis and resulted in improved saccharification [18–23], the time-consuming process of genetic perturbations in planta is still the bottleneck for the global understanding over lignin biosynthesis, as well as for the rational design of a promising biofuel plant.

Kinetic modeling provides opportunities to integrate the current knowledge of a given pathway in a dynamic manner [82, 83, 197]. It simulates the metabolism by a set of mass balance equations to track the changes in intermediate concentrations due to the differences between the synthesis and consumption rates, and reaction rates are formulated as functions of its substrates and catalyzing enzyme. Such a framework is able to incorporate individual enzymatic kinetics together into a dynamic model, as shown in several recent applications [91, 96, 97, 174, 198]. When the kinetic information for the pathway is incomplete, unknown parameters can be estimated by data fitting strategies with experimental inputs from metabolic and/or proteomic

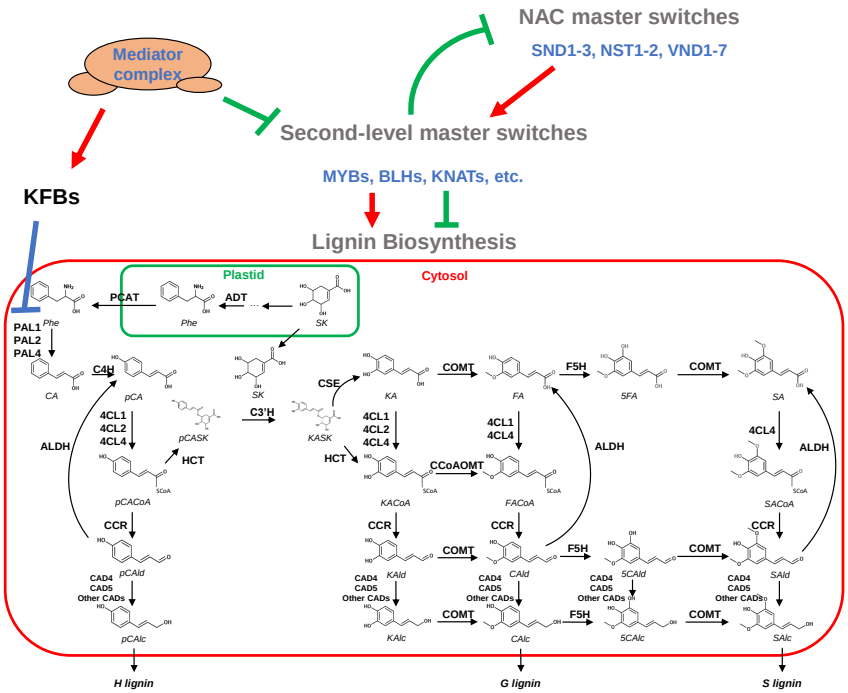

Fig. 4.1. Lignin biosynthesis in Arabidopsis. Key transcription factors and all metabolic steps involved in the lignin biosynthesis in Arabidopsis is shown. Transcription factors are colored in blue and structural genes in black. Intermediates and lignin products are shown in italic. Chemical reactions are represented by black arrows. Regulatory interactions for activation and inhibition are represented in red and green, respectively. Post-transcriptional regulation is shown in blue.

analysis [103, 109, 173, 199]. Such flexibility of kinetic modeling makes it possible to mechanistically integrate different types of information of the studied metabolism for a systems biological understanding. Further applications such as metabolic control analysis can utilize the constructed model to identify key steps in the pathway which control the metabolic flux, and this type of knowledge helps to determine targets for metabolic engineering [113]. In plants, kinetic modeling has been applied to primary carbon metabolism [96, 97, 105, 115, 198, 200, 201], as well as different kinds of secondary metabolism [91, 101–103, 109, 114, 116, 202].

Kinetic models for lignin biosynthesis were previously developed for *Populus trichocarpa* [92,174], *Medicago Sativa* L. [93], *Panicum virgatum* [94] and *Brachypodium distachyon* [95]. These models contributed to explore the dynamic responses of the pathway against different genetic perturbations [92–95, 174], and to propose and test the hypotheses of metabolic channeling *in silico* [93–95]. While kinetic models facilitate deeper understanding of lignin biosynthesis in plants, some limitations in model structures are present which leaves room for improvement. Previously, enzyme isoforms were simulated in a system-wise manner only in one study [174]. Generalized mass action models have difficulties in simulating all isoforms separately due to their empirical function nature. Therefore, without available kinetic information for each isoform, parameter identifiability is an issue for reliable estimation of all unknown parameters for a single metabolic step [92–95], and to propose and test the hypotheses of metabolic channeling *in silico* [93–95]. On the other hand, for those generalized mass action models, substrate competitions were incorporated into models only when it showed advantages for data fitting [93,94]. It is likely that many substrate competitions didnt play significant roles in their studied genetic lines, yet it limits the applications of models to the cases when those hidden factors were activated. [174] experimentally tested all possible substrate competitions by *in vitro* assays for all combinations of enzyme and inhibitor pairs in the pathway. However, we show that substrate competitions can be simulated in a more concise and systematic way in the current study, which bypasses possible error propagations from extra experiments. Moreover, none of above studies integrated the transcriptional regulation into the model, which is another major novelty in our work.

It is well known that a complex gene regulatory network (GRN) exists over lignin biosynthesis in plants [43, 49]; Fig. 4.1). NAM, ATAF1/2 and CUC2 (NAC) are master switches that were identified to be involved in the regulation of secondary cell wall synthesis in plants, including lignin biosynthesis [44–50]. They function together with second-level myeloblastosis (MYB) master switches to turn on the expression of downstream lignin biosynthetic genes [48, 49, 52, 203]. In addition to

this hierarchical regulatory structure, feedback control mechanisms from downstream transcription factors (TF) to the top-level master switches are also likely active in plants. The expression of SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 1(SND1) is inhibited by MYB4, MYB7 and MYB32 in Arabidopsis [52], while MYB4, MYB7 and MYB32 are all directly activated by MYB46 and MYB83 [53, 54], which are second-level master switches regulated by SND1 [53, 55, 56, 204]. The important regulatory role in lignin biosynthesis for the Mediator complex is also characterized in recent studies [22,58,205], and Mediator complex is likely regulating lignin biosynthesis via affecting the expression of MYB factors [58]. The Mediator complex also regulates the gene expression of Kelch repeat F-box (KFB) protein family, which are post-transcriptional regulators of the phenylalanine ammonia-lyase (PAL), thus controlling the first metabolic step in the lignin biosynthesis [59–61]. It is therefore important to understand the transcriptional level of lignin biosynthesis in addition to the metabolic level, if one aims to develop an in silico systems biological framework to predict the dynamic responses against various kinds of genetic perturbations.

Models for GRN of lignin biosynthesis has been previously developed for root xylem cells [140] and hypocotyledonous stem tissue [206] in Arabidopsis. [206] applied a mutual information based method to infer a two layered network between TFs and structural genes trained on microarray data. They showed their method outcompeting some other existing methods with higher accuracy in identifying positive regulatory mechanisms. Furthermore, they found it common to have multiple regulators functioning upstream of each structural gene in lignin biosynthesis. On the other hand, [140] have developed feedforward networks to describe the regulatory mechanisms against abiotic stress in plants. Models in [140] have a better model structure than ones in [206], since the former incorporated the interactions between TFs. While both of them successfully revealed the parts of lignin biosynthesis GRN in their studies, a full view of the transcriptional regulation is lacking due to the limited sizes of their models. Moreover, both studies only inferred the existence of a specific interaction without the knowledge

of regulatory strength, therefore the model cannot be applied in a dynamic way to predict the responses against genetic perturbations.

In the current study, we developed a dynamic model of lignin biosynthesis integrating the metabolic reactions with the transcriptional regulatory mechanisms. Due to the difficulties in parameterization of a large differential equation system, previous efforts for the hierarchical model development have mainly focused on the combination between flux balance analysis for metabolism and statistical models such as Boolean logic for transcriptional regulation [145–147]. While these models scale well on a genome scale, they sacrificed the mechanistic and dynamic properties. We overcame the challenges in parameterization by starting with a candidate set for gene interactions based on literature evidence, and simplified the system with first order kinetics for gene regulation. Model built in this way still preserves the dynamic nature and is well-compatible with our kinetic model for the metabolic network. The model was trained and validated with data from various genetic backgrounds, and its capability to uncover key changes in response to genetic perturbations was presented. Moreover, the model provides the opportunities to predict the consequences of all possible genetic manipulations for either structural genes or TFs in lignin biosynthesis, thus can be a great system for new hypothesis generation, as well as evaluation of our current understanding of lignin biosynthesis.

4.2 Materials and Methods

4.2.1 Plant materials and growth

Arabidopsis thaliana (Columbia-0) plants of different genotypes were grown in soil at 23 °C under 16/8-hour light/dark cycle. For consistency with [175], the basal 0.5-2 cm of the inflorescent stems were taken from four-week-old plants for RNA-seq and LC-MS/MS analysis. For the mutant genotypes used in the current study, *pal1* (SALK_096474C), *pal2* (GABI_692H09-025071), *4cl1* (WiscDsLox473B01), *cse-2* (SALK_023077), *cadC* (SALL_1265_A06), *cadD* (SALL_776_B06), *med5a* (SALK_011621)

and *med5b* (SALK_037472) were generated by T-DNA insertion and obtained from the Arabidopsis biological resource center; *fah1*, *ref2*, *ref3-2*, *ref3-3* and *ref8* were isolated from ethyl methanesulfonate-mutagenized populations with reduced sinapoyl malate contents [24, 207–209], which impacts the functioning of F5H, an enzyme in aliphatic glucosinolate biosynthesis, and C4H, respectively; *F5H* overexpresser (*C4H:F5H*) introduced a Arabidopsis *C4H* promoter driven *F5H* in the plant [18]; *SmF5H* introduced a Arabidopsis *C4H* promoter driven *Selaginella moellendorffii F5H* [210]. Higher order mutants were generated by crossing corresponding single mutants.

4.2.2 RNA-seq datasets

RNA-seq datasets for wild type, *pal1 pal2, ref3-2, ref3-3, 4cl1, cse-2, fah1, med5a/b ref8, med5a/b ref2, F5H*-overexpression, *ref8 fah1 SmF5H* and *med5a/b* genetic backgrounds in *Arabidopsis thaliana* Columbia-0 were taken from [211]. Four replicates were collected for each mutant genotype and eight for wild type. Briefly, RNA-seq analysis was performed on a single chip with an Illumina NovaSeq instrument at Purdue Genomics Core Facility. Fragments per kilobase of transcript per million mapped reads (FPKM) were generated with HISAT2 [212] and HTSeq-count programs [213] over TAIR10 genome build.

4.2.3 LC-MS/MS analysis

Metabolome datasets for wild type, *pal1 pal2, 4cl1, cse-2, fah1, med5a/b ref2, F5H*-overexpression and *med5a/b* genetic backgrounds in *Arabidopsis thaliana* Columbia-0 were obtained from the same plant materials as for the RNA-seq datasets [211]. Three replicates were collected for each genotype. All phenylpropanoid measurements in the datasets were obtained with LC-MS/MS as in [65]. Briefly, tissue samples were extracted and incubated in 75% methanol (v/v at 10 μL mg FW^{-1}) for 2 hours at 65 °C. The samples were centrifuged for 20 min at 16,000 x *g*, followed by drying

the supernatants in a speed-vac to concentrate the samples. The dried extracts were re-dissolved in 60 μL 50% methanol before separating by LC on a Zorbax Eclipse C8 column (150 mm 4.6 mm, 5 μm, Agilent Technologies, Santa Clara, CA), where the mobile phase is acetonitrile/H_2O/formic acid (9.8/2/0.2) buffered by 2.5 mM ammonium acetate (pH 5.6). Soluble metabolites were detected with a QTrap 5500 mass spectrometer (AB Sciex, Redwood City, CA) by multiple reaction monitoring in the negative ion mode. Calibration curves were applied for quantification, which were generated with standards of Phe, cinnamate, *p*-coumarate, caffeate, ferulate, sinapate, shikimate, *p*-coumaraldehyde, coniferaldehyde, sinapaldehyde, *p*-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol, *p*-coumaroyl shikimate, and caffeoyl shikimate.

4.2.4 Lignin analysis

Lignin content and monomer composition was analyzed as in [109]. Briefly, the cell wall residues were dried at room temperature for lignin analysis after being washed with 10 mL 70% ethanol (v/v) five times and once with acetone. Acetyl bromide lignin analysis was used to quantify total lignin content [177]. About 2 mg of cell wall residue was dissolved in 2.5 mL solution of acetyl bromide: glacial acetic acid (25:75) overnight at room temperature. Samples were mixed with 2.5 mL 2 M NaOH, 12 mL glacial acetic acid, and 0.5 mL 7.5 M hydroxylamine hydrochloride in a 50-mL volumetric flask with final volume brought to 50 mL with glacial acetic acid. Total lignin for each sample was measured with absorbance at 280 nm and calculated with the extinction coefficient of 23.29 g^{-1} L cm^{-1}. The rest of the cell wall residue sample was weighed and used for monomer composition analysis with DFRC/GC/FID/MS. Samples were dissolved overnight at room temperature in 2.5 mL of a solution of acetyl bromide: acetic acid (20:80) using 0.2 mg 4,4'-ethylidenebisphenol as internal standard. The products were dried and dissolved in 2 mL dioxane: acetic acid: water (50:40:10). Zinc dust (50 mg) was added to each sample and mixed well for 20 min. About 200 μL supernatant was then loaded onto a LC-Si solid phase extraction column

pre-conditioned with 95% ethanol and water and washed with 2.5 mL 25% ethanol. 2.5 mL 95% ethanol was used for sample elution, and then samples were dried before derivatization with 0.5 mL anhydrous acetic anhydride: pyridine (60:40) overnight. Acetylated products were then dried and dissolved in 200 μL dichloromethane. A 1 μL sample was analyzed on GC/FID for monomer composition and on GC/MS for labeling percentage of each monomer.

To determine *de novo* lignin deposition rate, the 0.5-2 cm stem sections from soil-grown plants were harvested at three days after bolting for wild type, *pal1 pal2* and *4cl1*. Total lignin content was quantified by acetyl bromide method [177]. Slopes were estimated with linear regression, which were used to determine the de novo lignin deposition rate (Supplementary Fig. C.1). The estimated deposition rate for wild type was similar with the previous report [109], therefore the previous value 3.44 nmol g FW^{-1} min^{-1} in wild type was used as the reference value, so that parameters from the previous model [109] can be directly transferred to the current study. *De novo* lignin deposition rates for *pal1 pal2* and *4cl1* were calculated by multiplying the reference rate with the relative ratios between *pal1 pal2/4cl1* and wild type obtained from Supplementary Fig. C.1. Deposition rates for *fah1* and *F5H*-overexpression genotypes were calculated by multiplying the reference rate with the relative ratios of mutant versus wild type. Deposition rates for *cse-2* and *med5a/b* genotypes were calculated by multiplying the reference rate with the relative ratios of mutant versus wild type from previous studies [21, 22]. Similarly, steady state fluxes towards each lignin subunit were calculated by multiplying the *de novo* deposition rate for each genotype (except for *pal1 pal2* and *4cl1*, which were measured from the current study) with the monomer compositions reported in previous studies [18, 21, 22, 210]. We do not have data for *med5 ref2* lignin content and monomer composition, therefore lignin subunit fluxes for this genotype was not constraint during model training.

4.2.5 PAL and 4CL enzyme assays

Stem tissue was harvested and frozen in liquid nitrogen. Crude protein was extracted from ground tissue with Tris-HCl buffer at pH of 7.8, and desalted on a gel filtration column (SephadexTM, G-50 fine, GE Healthcare). PAL and 4CL assays were conducted following the method in [214]. Each PAL assay contained 100 mM Tris-HCl buffer pH 7.5, 5 mM Phe, and 5 μL protein extract in a final volume of 50 μL. The reactions were incubated at 23 °C for 120 min. The 4CL assay contained 100 mM Tris-HCl buffer pH 7.5, 5 mM $MgCl_2$, 5 mM ATP, 1 mM *p*-coumarate, 0.3 mM CoA, and 2 μL protein extract in a final volume of 40 μL. Each reaction was incubated at 23 °C for 20 min. Assay products were quantified on HPLC with cinnamate and synthesized *p*-coumaroyl CoA as standards, respectively. Protein concentrations were measured with Bradford assay using bovine serum albumin as standard.

4.2.6 HCT enzyme assay

HCT assay was coupled with 4CL reaction. Purified (22 ng) 4CL enzyme expressed in *Escherichia coli* was used to synthesize *p*-coumaryl CoA or caffeoyl CoA substrate at 23 °C for 3 hours before the HCT assay. HCT reaction was initiated by adding protein extract and lasted for 20 min at 23 °C. The assay mixture contained 100 mM Tris-HCl buffer pH 7.5, 5 mM $MgCl_2$, 2.4 mM ATP, 2.4 mM CoA, 2 mM *p*-coumarate or caffeate, 7.5 mM shikimate, 2 μL of 4CL enzyme, 8 μL protein extract in a final volume of 40 μL. Assay product *p*-coumaroyl shikimate or caffeoyl shikimate was quantified on HPLC. Protein concentrations were measured with Bradford assay using bovine serum albumin as standard.

4.2.7 CCR enzyme assay

CCR assay was coupled with 4CL reaction. Purified 4CL (22 ng) enzyme expressed in *Escherichia coli* was used to synthesize *p*-coumaryl CoA substrate at 23 °C for 3

hours before the CCR assay. One unit glucose 6-phosphate dehydrogenase was used to synthesize NADPH in a mixture of 1 mM $NADP^+$, 10 mM glucose 6-phosphate at 30 °C for 5 min. This NADPH synthesis mixture (12 μL) was added to each CCR assay mixture. The CCR assay was initiated by adding protein extract and incubated for 20 min at 23 °C. The assay contained 100 mM Tris-HCl buffer pH 7.5, 5 mM $MgCl_2$, 1 mM ATP, 0.15 mM CoA, 0.1 mM *p*-coumarate, 0.2 mM NADPH, 1 μL of 4CL enzyme, 4 μL protein extract in a volume of 60 μL. The product *p*-coumaroyl aldehyde was quantified on HPLC. Protein concentrations were measured with Bradford assay using bovine serum albumin as standard.

4.2.8 Gene regulatory network reconstruction

In total 438 candidate gene regulatory interactions were collected from literature, specifically for lignin biosynthesis in Arabidopsis. These interactions were used to define an initial gene regulatory network of lignin biosynthesis in Arabidopsis. Mathematically, the network was formulated with linear differential equations, where a mass balance equation is formulated as:

$$\frac{d[G_i]}{dt} = Const_i + \sum_j K_{reg,j,i}[P_j] - K_{decay,i}[G_i] \tag{4.1}$$

Where $[G_i]$ represents the mRNA amount of the gene involved in the lignin biosynthesis; $Const_i$ is the constitutive synthesis rate for $[G_i]$, which summarizes the transcriptional driving forces in addition to the regulatory mechanisms in lignin biosynthesis; $K_{reg,j,i}$ is the regulatory strength exerted by the transcription factor $[P_j]$; $K_{decay,i}$ is the mRNA decay rate for $[G_i]$. $K_{decay,i}$ values were collected from [215], with missing values filled with the median decay rate. In addition, observations from [216] suggested that although there is no strong correlation between transcriptome and proteome data, the correlation between a given genes expression and its associated protein amount seems preserved along different conditions. These observations justified

the assumed linearity between mRNA and protein amounts corresponding to each gene, converted with a transfer coefficient α_j:

$$[P_j] = \alpha_j[G_j] \tag{4.2}$$

Therefore equation 4.1 can be adjusted to exclude the needs to simulate protein amounts for transcription factors explicitly:

$$\frac{d[G_i]}{dt} = Const_i + \sum_j K'_{reg,j,i}[G_j] - K_{decay,i}[G_i] \tag{4.3}$$

Where $K'_{reg,j,i}$ is the apparent regulatory strength for $[G_j]$.

FPKM values were used in the model to represent the gene expression levels in each genetic background. For the simulation of mutant generated by T-DNA insertion, the synthesis rates of the corresponding gene (both $Const_i$ and $K'_{reg,j,i}$) were multiplied with a correction factor of zero to mimic the mutation event. For the simulation of *F5H*-overexpression line, the synthesis rates of F5H were multiplied with a correction factor of 61.5, which was obtained from the F5H ratio between *fah1* and wild type lines. To estimate $Const_i$ and $K'_{reg,j,i}$ values in the model, RNA-seq datasets from wild type, *pal1 pal2, ref3-2, ref3-3, 4cl1, cse-2, fah1, ref8 fah1 SmF5H, med5a/b ref8* and *med5a/b ref2* genetic backgrounds were applied for model training. Parameters were estimated by optimizing the following objective function:

$$\hat{sse}_{GRN} = \sum_{i=1}^{N}\sum_{j=1}^{F_i}\left(\frac{[G_{i,j}^{predicted}] - [G_{i,j}^{measured}]}{s_{i,j}}\right)^2 \tag{4.4}$$

With F_i measurements of FPKM for genes in the network for each of N datasets, $s_{i,j}$ is the standard deviation estimated from $n = 4$ for each condition ($n = 8$ for wild type). Note that the equation 4.3 reduces to a linear equation when the system reaches a steady state, therefore the optimization problem was solved by iteratively calculating the solution for the linear equations for a given parameter set. The *lsqnonlin* function in MATLAB (R2013a, Mathworks, Inc) was then applied to determine the changing directions for the parameters for each iteration. The final solution was achieved when the algorithm stopped, and its robustness was verified with *ode15* integration independently starting from an initial point where all mRNA amounts were zeros.

4.2.9 Kinetic modeling

The kinetic model of lignin biosynthesis in Arabidopsis consists of reactions starting from plastidial Phe synthesis, and ending up with the production of five aldehydes and five alcohols which can be incorporated into the lignin polymer. First order kinetics are assumed for the lignin incorporation for each aldehyde/alcohol. In addition, all acids are assumed to have sink reactions with first order kinetics. Mass balances of pathway intermediates are simulated with the equation below (See Supplementary Material D.1 for detailed mass balances):

$$\frac{dC}{dt} = Sv \tag{4.5}$$

where C is the vector of metabolite concentrations, v is the vector of reaction velocities, and S is the stoichiometric matrix for the pathway. Michaelis-Menten kinetics are applied to simulate the reaction rates. Since several enzymes are capable of utilizing multiple substrates, those substrates act as competitive inhibitors towards each other for common catalytic sites [217]. In general, the rate equations take the form as shown below (See Supplementary Material D.2 for detailed rate equations for the kinetic model):

$$v_{E,A} = V^{A}_{max,E} \frac{\frac{[A]}{K^{E}_{m,A}}}{1 + \frac{[A]}{K^{E}_{m,A}} + \frac{[B]}{K^{E}_{m,B}}} \tag{4.6}$$

where $v_{E,A}$ is the rate of the reaction catalyzed by the enzyme E and converts A to the downstream product; $[A]$ and $[B]$ are the concentrations of competing substrates; $V^{A}_{max,E}$ is the maximal capacity for E catalyzing substrate A, with the binding affinity of $K^{E}_{m,A}$ (Similarly, $K^{E}_{m,B}$ is the binding affinity of B). Parameters for Phe biosynthesis including ADT and PCAT kinetics were taken from a previous phenylpropanoid model [109]. For the rest of the pathway, most K_m values for enzymes involved in the lignin biosynthesis in Arabidopsis were collected from literature (Supplementary Table C.1), with the exceptions for all parameters for HCT, $K^{4CL}_{m,\ CoA}$ for 4CL isoforms and $K^{CCoAOMT}_{m,\ KACoA}$ for CCoAOMT. To estimate the cytosol volume for unit conversion for K_m, water content in 0.5-2 cm basal stem was first measured

to be 0.96 ± 0.01 mL g FW^{-1}. The fraction of active lignifying cells was estimated to be 34.2% from a microscopic cross-section of a typical Arabidopsis basal stem, by excluding pith and epidermis regions. By applying the estimation of cellular fraction of cytoplasm from [181], the final converting factor was calculated to be 0.022 mL g FW^{-1}. k_{cat} or V_{max} values for PALs, 4CLs, CADs, CCR, COMT, F5H and ALDH were also collected from literature to capture different catalytic efficiencies for competing isoforms/substrates (Supplementary Table C.1). The catalytic efficiencies for Arabidopsis HCT are taken as free variables in the model as there were no literature values. Since HCT catalyzes two reversible reactions, Haldane relationship [217] was applied, which reduced the total number of unknown kinetic parameters for HCT to seven. Specifically, kinetic parameters of HCT need to satisfy the following relationships:

$$K_{eq,\ HCT}^{pCACoA} = \frac{k_{cat,\ HCT}^{pCACoA} K_{m,\ pCASK}^{HCT} K_{m,\ CoA}^{HCT}}{k_{cat,\ HCT}^{pCASK} K_{m,\ pCACoA}^{HCT} K_{m,\ SK}^{HCT}} \tag{4.7}$$

$$K_{eq,\ HCT}^{KASK} = \frac{k_{cat,\ HCT}^{KASK} K_{m,\ KACoA}^{HCT} K_{m,\ SK}^{HCT}}{k_{cat,\ HCT}^{KACoA} K_{m,\ KASK}^{HCT} K_{m,\ CoA}^{HCT}} \tag{4.8}$$

$K_{eq,\ HCT}^{pCACoA}$ and $K_{eq,\ HCT}^{KASK}$ were estimated to be 0.411 and 2.334 with eQuilibrator [218, 219] assuming ionic strength of 0.33, pH of 7.4 and temperature of 25 °C. The $k_{cat,\ HCT}^{pCACoA}$ was set as 1 s^{-1} to further reduce one variable. Note that since all k_{cat} values were multiplied with corresponding estimated enzyme amounts to obtain V_{max} values in the model, only relative ratios of k_{cat} for competing isoforms/substrates were important. In each genotype simulation, the enzyme amount ratios for isoforms in a common enzyme family were assumed to be equal with the corresponding mRNA expression ratios under corresponding genetic condition, so that each enzyme family can be simulated by only one free variable to represent one isoforms concentration, with the rest of isoform levels in the same family derived from mRNA ratios. Post-transcriptional regulation of KFBs to PALs was simulated with a linear equation:

$$[P_{PAL_i}] = \alpha_{PAL}[G_{PAL_i}] - K_{KFB}([G_{KFB01}]+[G_{KFB20}]+[G_{KFB39}]+[G_{KFB50}]),\ \ i = 1, 2, 4 \tag{4.9}$$

There are 12 parameters for enzyme abundances, 9 parameters for unknown kinetic parameters, three for initial concentrations of Phe, CoA and shikimate, two for first order kinetic coefficients for acid and lignin sinks, one for the KFB regulatory factor, summing up to 27 free parameters to be estimated for the kinetic model. The whole differential equation system was integrated with *ode15* function in MATLAB. For each simulation, modeling outputs were obtained until the system reached the steady state, starting from an initial point where all intermediate pools being zeros, except for CoA and shikimate.

The model was trained with datasets from wild type, *pal1 pal2, 4cl1, cse-2, fah1* and *med5a/b ref2* genetic backgrounds with in total 176 measurements. Lignin subunit fluxes and metabolite concentrations were available for each genetic line, and wild type has additional measurements of maximal capacities for PAL, 4CL, HCT and CCR. While all these measurements were applied for parameter estimation, a weighted objective function was applied:

$$\begin{aligned} s\hat{s}e_{kinetic\_model} = &\sum_{i=1}^{Q}\Big[10\sum_{j=1}^{L}\Big(\frac{Y_{lig,i,j}^{predicted}-Y_{lig,i,j}^{measured}}{s_{lig,i,j}}\Big)^2+\sum_{m=1}^{M_i}\Big(\frac{Y_{met,i,m}^{predicted}-Y_{met,i,m}^{measured}}{s_{met,i,m}}\Big)^2\Big] \\ &+\sum_{r=1}^{R}\Big(\frac{Y_{vmax,wt,r}^{predicted}-Y_{vmax,wt,r}^{measured}}{s_{vmax,wt,r}}\Big)^2 \end{aligned} \tag{4.10}$$

With M_i metabolite concentrations, L lignin subunit fluxes for each of Q datasets, and R maximal capacity measurements for wild type condition, $s_{i,j}$ is the standard deviation estimated from $n = 3$ for each condition. The training procedure was the same as [109]. Briefly, *lsqnonlin* in MATLAB was used with multiple starts to generate a starting point for Markov Chain Monte Carlo (MCMC) sampling [108], and MCMC batches each with 100, 000 samples were generated until no further reduction in $s\hat{s}e_{kinetic\_model}$ was observed for a new batch. The parameters with the lowest $s\hat{s}e_{kinetic\_model}$ were then selected as the point estimations, and the second to last batch of MCMC samples were used to estimate the variations for parameters and model outputs. Samples for variance analysis were selected with an interval size of 100 to reduce the autoregression effect in MCMC chains.

4.3 Results

4.3.1 Transcriptional/post-transcriptional regulatory interactions for lignin biosynthesis in Arabidopsis were compiled from the literature

A large number of studies has been carried out to identify transcriptional/post-transcriptional regulatory mechanisms in lignin biosynthesis in Arabidopsis thaliana [44–46,48,49,52,140,203,206], which has resulted in a much more comprehensive view of the regulation of lignin biosynthesis compared with other plant species [51]. We aimed at building an evidence-derived GRN at first by collecting all interactions reported so far in Arabidopsis. Our focus has limited on the experimental evidence which provides the information of interaction type (activation/inhibition). This constraint reduced our literature search to the evidence from steroid receptor-based inducible system [220], glucocorticoid receptor-mediated post-translational inducible system [46], *in vivo* chromatin immunoprecipitation [55], electrophoretic mobility shift assay [55], transfection assay [221], histochemical β-glucuronidase staining assay [222] as well as significant expression changes in TF perturbed genetic lines. The former four experiments provide the evidence for direct binding of TF to the target genes promoter region, and the later three only provide indirect evidence for regulation without evaluation of direct binding. In total, we have collected 487 transcriptional/post-transcriptional regulatory interactions between 89 genes reported in Arabidopsis (Supplementary Table C.3). Since the purpose of the interaction collection was to construct a GRN of genes that regulate structural gene (enzymes) expression in lignin biosynthesis, we further removed TFs without any downstream targets and their associated upstream interactions, ending up with 438 transcriptional regulatory interactions and 12 post-transcriptional regulatory interactions between 72 genes (Supplementary Fig. C.2). Among 450 collected interactions, 23.6% (106 out of 450) of them have evidence for direct binding. We listed the top ten TFs (Table 4.1) and structural genes (Table 4.2) with highest number of reported interactions. MYB46 is the hub node in the network, with 10 TF regulators and 30 gene targets reported so

Table 4.1.
Top 10 TFs with highest number of interactions reported.

		Upstream[a]		Downstream[b]		
Gene Name	Locus	Direct[c]	Indirect[d]	Direct	Indirect	Total
MYB46	AT5G12870	5	5	21	9	40
MYB63	AT1G79180	2	14	1	17	34
SND1	AT1G32770	1	6	8	16	31
KNAT7	AT1G62990	8	7	1	12	28
MYB83	AT3G08500	5	5	12	6	28
NST1	AT2G46770	2	1	5	18	26
MYB103	AT1G63910	5	12	0	8	25
MED5a	AT3G23590	0	0	0	23	23
MED5b	AT2G48110	0	0	0	23	23
MYB58	AT1G16490	2	10	9	1	22

[a]Number of regulators;

[b]Number of targets regulated by the gene;

[c]Number of interaction with direct binding evidence;

[d]Number of interaction without direct binding evidence.

far. MYB83, the ortholog of MYB46, also ranked in top ten list. Besides MYB factors and NAC master switches, KNAT7 and Mediator subunits MED5a and MED5b also ranked in the top ten list (Table 4.1). For structural genes in lignin biosynthesis, most findings of regulatory interactions have been reported in the upstream part of the pathway, as PAL1 and 4CL1 ranked the top two in the list (Table 4.2).

4.3.2 Gene regulatory network was reconstructed from the collected interactions with linear differential equations

A linear differential equation system was developed to represent the dynamic interactions within the GRN for lignin biosynthesis in Arabidopsis, this consists of

Table 4.2.
Top 10 Structural genes with highest number of interactions reported.

		Upstream[a]		
Gene Name	Locus	Direct[b]	Indirect[c]	Total
4CL1	AT1G51680	3	26	29
PAL1	AT2G37040	9	19	28
CCoAOMT	AT4G34050	5	22	27
HCT	AT5G48930	3	20	23
PAL4	AT3G10340	5	12	17
C4H	AT2G30490	4	12	16
C3'H	AT2G40890	2	13	15
F5H	AT4G36220	1	12	13
CCR	AT1G15950	2	10	12
COMT	AT5G54160	3	9	12

[a]Number of regulators;

[b]Number of interaction with direct binding evidence;

[c]Number of interaction without direct binding evidence.

438 transcriptional regulatory interactions from the literature (See Materials and Methods for more details). Only the dynamics of genes were simulated in the GRN, with TF protein amounts implicitly incorporated into the apparent K'_{reg} values for conciseness. To parameterize the GRN model, we applied RNA-seq datasets from wild type, *pal1 pal2, ref3-2, ref3-3, 4cl1, cse-2, fah1, ref8 fah1 SmF5H, med5a/b ref8* and *med5a/b ref2* genetic backgrounds for model training. There are in total 511 parameters to be estimated from 710 measurements. The model was further validated with data from *cadC cadD, F5H*-overexpression and *med5a/b* genetic backgrounds, and showed good matches with experimental measurements in both training and validation datasets (Fig. 4.2a). To evaluate the number of active interactions predicted by the

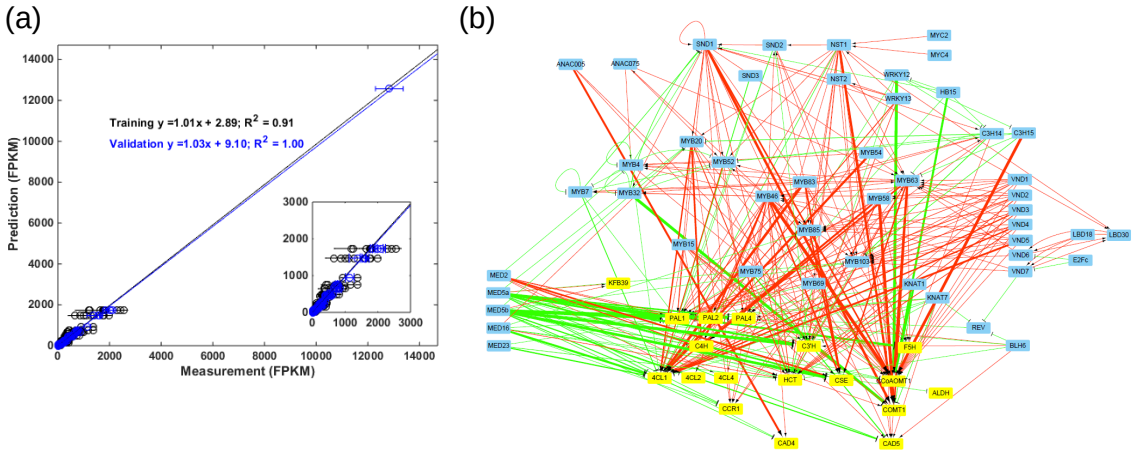

Fig. 4.2. GRN reconstruction for lignin biosynthesis in Arabidopsis. (a) Comparison between models prediction versus experimental measurements for training and validation datasets. Adjusted coefficients of determination were calculated to evaluate the models performance for both training and validation datasets. (b) The GRN for lignin biosynthesis in Arabidopsis after parameter rounding. Blue nodes represent TFs and yellow nodes represent structural genes. Arrows represent the regulatory interactions between a TF and its target gene. Activation is represented by red color and inhibition by green. The thickness of arrows are proportional to the product of the regulatory coefficient and the corresponding TFs expression in wild type. The plot was generated with Cytoscape 3.7.0.

model, parameter rounding was implemented by setting values smaller than 0.00001 to zero. We verified the gene expression predicted by the model after parameter rounding only differed at most 3% from the one predicted by the model without any rounding, and the rounded model has 320 interactions in total with non-zero regulatory coefficients (Fig. 4.2b). To investigate the regulation exerted by the Mediator complex subunits MED5a/b, the subnetwork of MED5a/b was extracted from the GRN by selecting all TFs and structural genes regulated by MED5a/b or their downstream TFs (Fig. 4.3). Interestingly, distal interactions between MED5a/b and SND1 were present in the subnetwork, which were connected with a series of MYB factors. On the

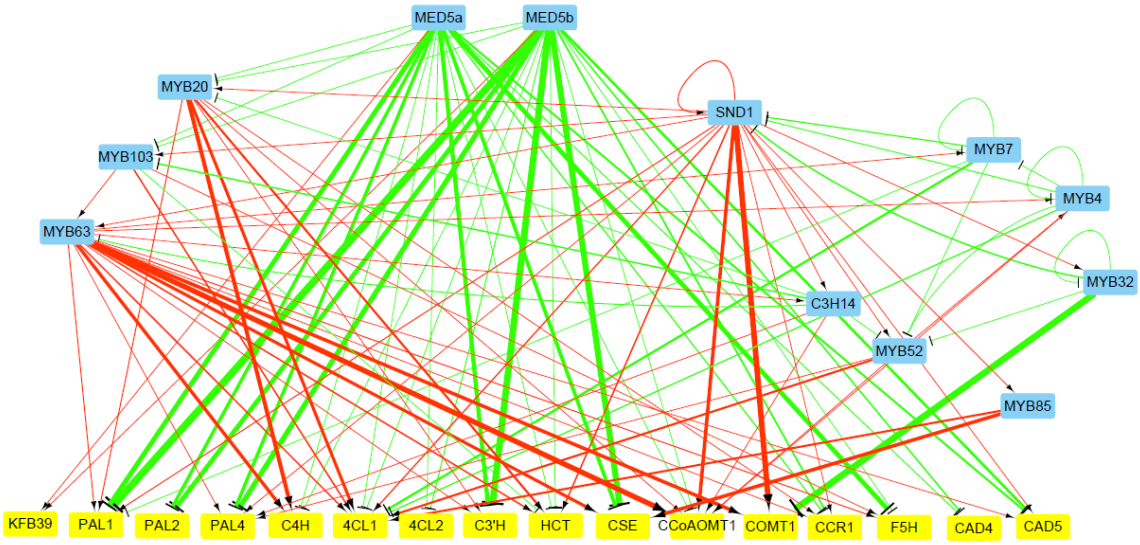

Fig. 4.3. Subnetwork of med5a/b for lignin biosynthesis in Arabidopsis. Blue nodes represent TFs and yellow nodes represent structural genes. Arrows represent the regulatory interactions between a TF and its target gene. Activation is represented by red color and inhibition by green. The thickness of arrows are proportional to the product of the regulatory coefficient and the corresponding TFs expression in wild type. The plot was generated with Cytoscape 3.7.0.

other hand, structural genes in lignin biosynthesis seem to be regulated by MED5a/b mainly with yet unknown mechanisms other than MYB factors, as the arrows between MED5a/b and structural genes were obviously thicker than ones between MED5a/b and MYB20/MYB103 in general (Fig. 4.3).

4.3.3 Kinetic modeling of lignin biosynthesis in Arabidopsis was constructed and parameterized with a weighted objective function

To mathematically describe the metabolic dynamics for lignin biosynthesis in Arabidopsis, a kinetic model of the pathway was developed, which includes all known isoforms and substrate competitions in the metabolic network (See Fig. 4.1 and

Supplementary Material D.2 for more details). The information of all k_{cat}/K_m values in previous literature were collected to parameterize the model (Supplementary Table C.1). Since no literature reference values were available for the K_m value of CoA for 4CLs, the K_m value of caffeoyl CoA for CCoAOMT, as well as k_{cat} and K_m values of all substrates for HCT in Arabidopsis, they were made to be free parameters in the model. To reduce the burden of the model parameterization and also for parameter consistencies, we further assumed that all 4CL isoforms share a common K_m value for CoA; and Haldane relationship [217] was applied to reduce the number of free parameters for HCT step, with estimated equilibrium constants [218, 219]. [216] observed that the correlation between a given genes expression and its associated protein amount seems preserved along different conditions, which justifies the linearity assumption between the enzyme amount and the mRNA level for the associated gene. Therefore, the V_{max} values in the model were calculated by the product between enzyme amounts and k_{cat} values, with the former values further derived from the products between the mRNA levels and the mRNA-protein transfer coefficients. Each enzyme family shared a common transfer coefficient in the model, so that the relative ratios of each isoform were proportional to their gene expression ratios, respectively. For the end product formation, since monolignols other than H, G or S subunits were observed in lignin [23, 32, 195, 196], we assume the lignin polymerization to be a flexible process, so that all aldehydes and alcohols can be incorporated into the polymer in the model. First order kinetics were assumed for the lignin incorporation for each aldehyde/alcohol. This simplification is due to the lack of mechanistic understanding for the polymerization process. Thus, mass action kinetics was assumed. To allow for new steady states when some upstream enzymes are blocked, sink reactions were added for all acids in the pathway. In fact, the accumulation of acid derivatives has been frequently reported in different mutant lines of lignin biosynthesis when lignification was impaired [223]. Free parameters were also set for initial concentrations of Phe, CoA and shikimate, implicitly assuming no net synthesis exists for CoA and shikimate during the modeling periods. For the simulation of each genetic condition, enzyme amounts were adjusted by the mRNA

levels, and the model was integrated with the initial condition in which the rest of metabolite concentrations being zero. The integration was stopped once the steady state was reached, so that the whole simulation was to represent the activation and establishment of a stable lignification process in plants.

Free parameters were estimated by training the model with data from wild type, *pal1 pal2, 4cl1, cse-2, fah1* and *med5a/b ref2* genetic backgrounds. The parameterized model was validated with data from *F5H*-overexpression and *med5a/b* lines (Fig. 4.4). Each genetic condition has data for phenylpropanoid concentrations and lignin subunit fluxes, and wild type condition has Vmax measurements for PAL, 4CL, CCR and HCT as well. The objective function weighted a factor of 10 for lignin subunit fluxes. Such weighting treatment took into consideration of the fact that the metabolome measurements are indeed the summation of multiple subcellular pools [80, 109, 167, 168], which not necessarily represent the cytosolic condition. Therefore, instead of attempting to fit all measurements equally well, we encouraged the model to fit the lignin subunits data better since the main purpose of the modeling is to reliably predict phenotypes of lignification in different genetic backgrounds. While the choice of the weighting factor can be further tuned, the model fit in this way had a reasonable prediction for lignin subunit fluxes for both training and validation datasets (Fig. 4.4).

4.3.4 A hierarchical model was developed to combine the GRN and the kinetic model of lignin biosynthesis

The kinetic model presented in the previous section requires the input of mRNA levels from the genetic line to be simulated, which limits the potential applications of the model in the situation where gene expression measurements are missing. The GRN developed in the current study provides a reasonable prediction for gene expression in response to different genetic perturbations (Fig. 4.2a), thus is a great system to generate gene expression inputs for the kinetic model. To test the idea of this hierarchical modeling, the kinetic model was trained and validated with the same

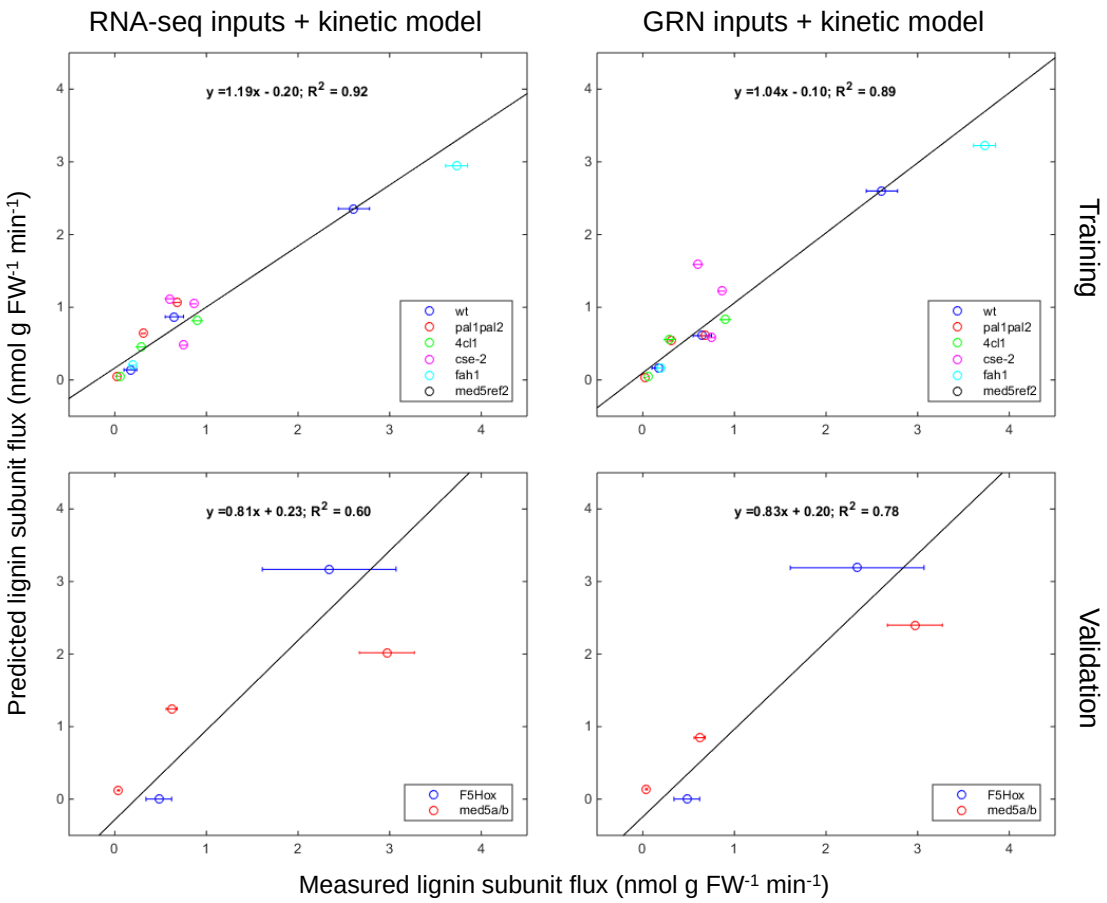

Fig. 4.4. Data fitting comparison between kinetic models with gene expression inputs from RNA-seq datasets or the GRN. The performance was compared between the kinetic model with gene expression inputs from RNA-seq datasets (left column) and the model with inputs from the GRN (right column). Both models were trained with the same datasets from six genetic backgrounds (first row), and validated with data two independent genetic backgrounds (second row). Note that the GRN was trained without using any information from *F5H*-overexpression and *med5a/b* lines, which ensures a fair comparison. Adjusted coefficients of determination were calculated to evaluate the models performance for both training and validation datasets.

datasets as presented in the previous section, with the only difference being the gene expression inputs from the GRNs prediction (Fig. 4.4). Note that the GRN was trained

without any data from *F5H*-overexpression and *med5a/b* backgrounds, so that the fairness of the comparison was kept. Overall both models have similar performance for a reasonable prediction of lignin subunit fluxes. The hierarchical model ended up with a slightly worse value of the adjusted coefficient of determination for training data, which was largely due to the overestimation of the H subunit flux in *cse-2* background. Other than that, lignin subunit fluxes predicted by the hierarchical model had good matches with measurements, just as the model with RNA-seq inputs for training data. The hierarchical model even had a better performance for the flux prediction in *med5a/b* background, which is in the testing datasets.

While a weighted objective function was applied for model training to fit the lignin data particularly, the hierarchical model had accurate predictions for measured maximal capacities in wild type (Fig. 4.5a). PAL, 4CL, HCT and CCR all had significantly higher capacities compared with lignin deposition rate 3.44 nmol g FW^{-1} min^{-1} in wild type (Supplementary Fig. C.1 and Fig. 4.5a), indicating the pathway is likely running under a substrate limiting condition. Predicted maximal activities for other enzymes were consistent with the substrate limiting hypothesis, except for F5H and ALDH (Fig. 4.5b-d). On the other hand, the model seems to have difficulties in fitting measurements for metabolite concentrations (Supplementary Fig. C.3). In fact, most metabolite amounts were underestimated by the model, except for Phe, coniferyl alcohol, and some acids. Such deviations either indicate the limitations in the current model structure, or perhaps implying that the measured metabolite pools are in an additional location besides lignifying cells cytoplasm. Consistent with the latter hypothesis, multiple phenylpropanoids have been found in vacuole [80, 167, 168], and sinapate is expected to accumulate in epidermal cells, as it is the precursor of sinapate esters for UV protection in plants [24, 207–209, 224]. The full list of estimated parameters for the hierarchical model can be found in Supplementary Table C.2.

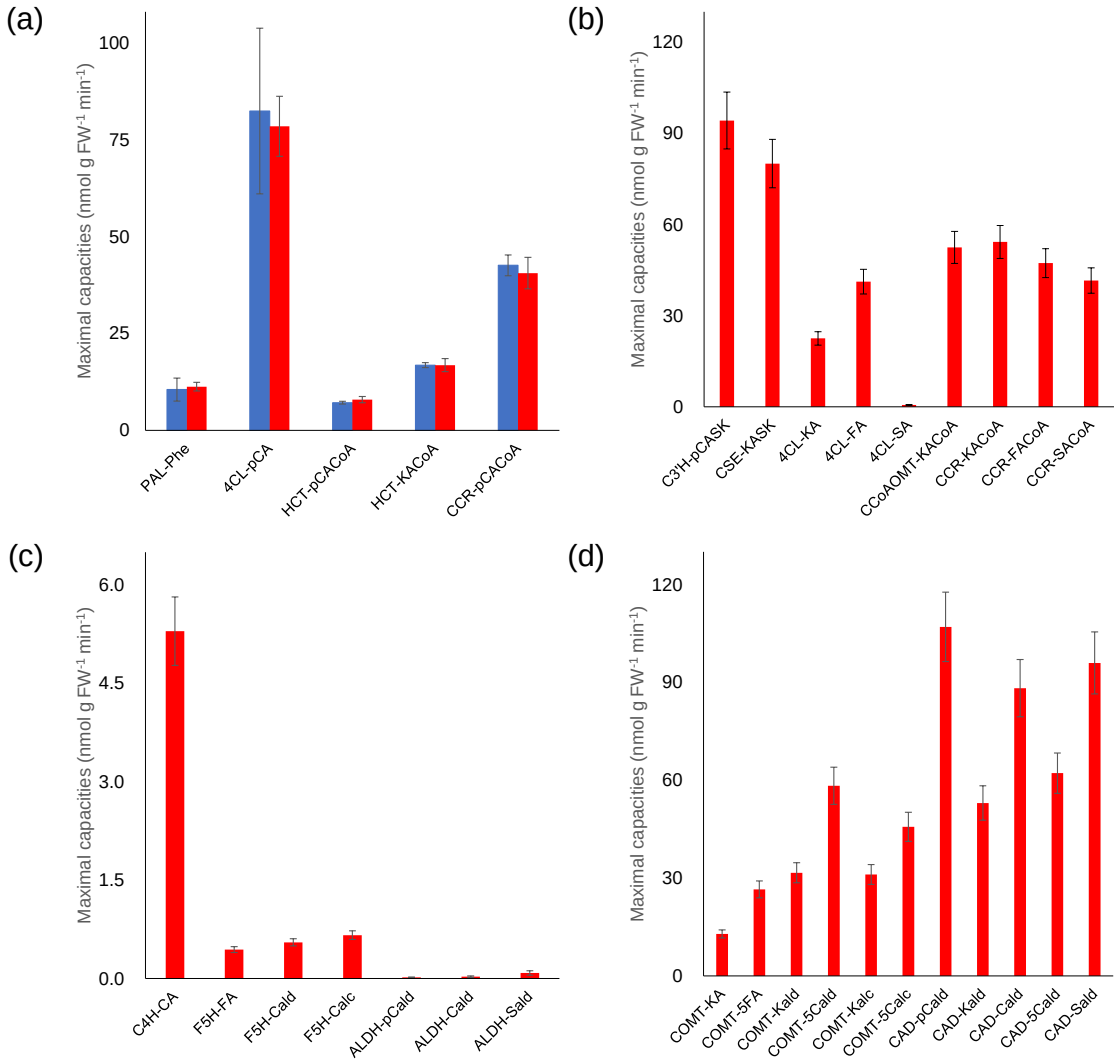

Fig. 4.5. Maximal enzyme capacities in the hierarchical model. Maximal enzyme capacities were calculated from the hierarchical model under wild type condition. Capacities for PAL, 4CL and CAD are the summation of all isoforms. Bar labels are represented as enzyme name substrate to differentiate the competing reactions. Bars are represented as point estimation ± standard deviation, with the latter calculated from 1,000 MCMC samples. (a) Maximal capacities for PAL, 4CL, HCT and CCR were compared between measurements (blue bars) and predictions (red bars). (b-d) Other maximal capacities from the hierarchical model.

4.3.5 Flux maps for different genetic backgrounds were derived from the hierarchical model

When a kinetic model reaches a steady state, metabolic fluxes in the model will be constant over time, so that the flux map in this case represents the stable carbon allocation at branch points and to different end products. We generated a lignin biosynthesis flux map in wild type Arabidopsis to investigate the carbon allocation in the pathway (Fig. 4.6). The model predicted 75% of entering flux ended up as G lignin, 18% for S lignin and 5% for H lignin. A minor output flux came out from coniferaldehyde as well. Interestingly, the model suggested that the majority of flux going to G and S lignin synthesis was through CSE instead of HCT, which provided a quantitative estimation of the contributions from these two alternative routes (Fig. 4.6). This is consistent with the observations that CSE is important for the biosynthesis of G and S lignin [21]. And the reduced lignin content in *HCT*-RNAi plants [225] is likely due to the perturbations to the first HCT step. 4CLs sustained a large flux not only for *p*-coumaroyl CoA, but also for caffeoyl CoA production (Fig. 4.6). This might explain the reason why plants invest a large amount of enzyme for 4CL steps (Fig. 4.5), as the fluxes going through 4CLs were at least two-fold compared with other enzymes in the model (Fig. 4.6). S lignin was predicted to be synthesized mainly from coniferyl alcohol, and F5H was functioning close to saturation to compete for coniferyl alcohol from G lignin synthesis (91% of its maximal capacity, see Fig. 4.5c and Fig. 4.6). Being the only enzyme working at full capacity, the reaction rate catalyzed by F5H for coniferyl alcohol consumption was much less sensitive to the reduced upstream fluxes (Supplementary Fig. C.4 – C.6). Therefore S/G ratios were increased in response to the upstream flux reduction, as G lignin synthesis was much more sensitive to those perturbations than S lignin in the model. This provides an explanation for the increased S/G ratios in various genotypes with impaired upstream enzymes [21, 31, 35, 209].

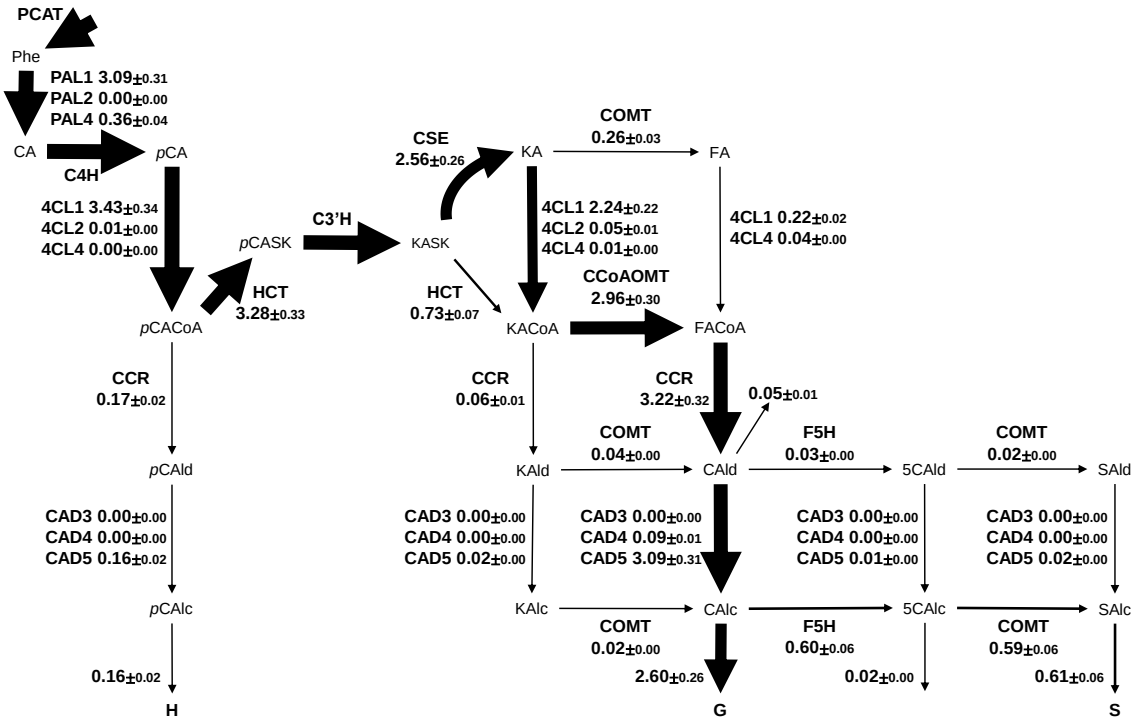

Fig. 4.6. Flux map of lignin biosynthesis in wild type Arabidopsis. The map was generated from the hierarchical model once it reached a steady state. The thickness of arrows are proportional to the reaction rates. Numbers next to the arrows are shown as point estimation ± standard deviation, with the latter calculated from 1,000 MCMC samples (Unit: nmol g FW^{-1} min^{-1}).

The most significant reduction in lignin deposition was observed for *pal1 pal2* and *4cl1* simulation (Supplementary Fig. C.4, C.5), with only 35% and 43% of wild type flux, respectively. However, model applied different strategies to deal with the perturbations to reach to a new steady state. When PAL1 and PAL2 were deleted, hyperaccumulation of Phe in the cytosol feedback inhibited the plastidial export, which activated the feedback inhibition of plastidial Phe to its own synthesis. This led to an overall reduced flux entering the pathway (Supplementary Fig. C.4). On the other hand, when 4CL1 was blocked, acid sinks were activated for *p*-coumarate and ferulate to dissipate extra carbon fluxes due to the saturations of 4CL2 and 4CL4.

These flux balancing mechanisms were missing from the previous kinetic model in Poplar [174], and we show that the presence of these mechanisms not only improved the models robustness, but also enabled the model to reproduce the observations in *pal1 pal2* [35, 80] and *4cl1* [223] plants.

PAL1, 4CL1 and CAD5 carried the majority of flux going through the metabolic steps with isoforms. While other isoforms didnt play significant roles in wild type simulation, they were activated when major isoforms were blocked (Supplementary Fig. C.4, C.5). PAL2 was predicted to be completely inhibited by KFBs in wild type (Fig. 4.6), while this is likely due to the model artifacts, activation of PAL2 was observed in the *med5a/b* simulation (Supplementary Fig. C.7). It is unclear whether these results are of physiological relevance at this point. On the other hand, when 4CL1 was knocked out, 4CL2 and 4CL4 were shown to function at different steps to compensate for 4CL1 loss (Supplementary Fig. C.5). 4CL2 was the major isoform responsible for *p*-coumaroyl CoA synthesis, and 4CL4 directed flux from CSE branch to the G and S lignin biosynthesis. *p*-Coumarate and ferulate outcompeted caffeate for 4CLs occupation, thus caffeate consumption by 4CLs did not carry significant flux in *4cl1* simulation (Supplementary Fig. C.5).

4.3.6 Hierarchical model can be used to generate *in silico* predictions for various genetic perturbations in lignin biosynthesis

Lignin amount and composition are major factors contributing to cell wall recalcitrance [187, 194], and it has been shown that a lower lignin content and a higher S/G ratio are correlated with the improved saccharification yield [20]. Therefore plants, with reduced lignin content or adjusted monolignol composition are desired. To evaluate the models capability in predicting these lignin phenotypes in different genetic backgrounds, we generated the gene expression levels with the GRN by adjusting the synthesis of each gene in the network with a factor of zero, 0.5 and 2 to simulate the effects of knockout, knockdown and overexpression (Fig. 4.7). Steady

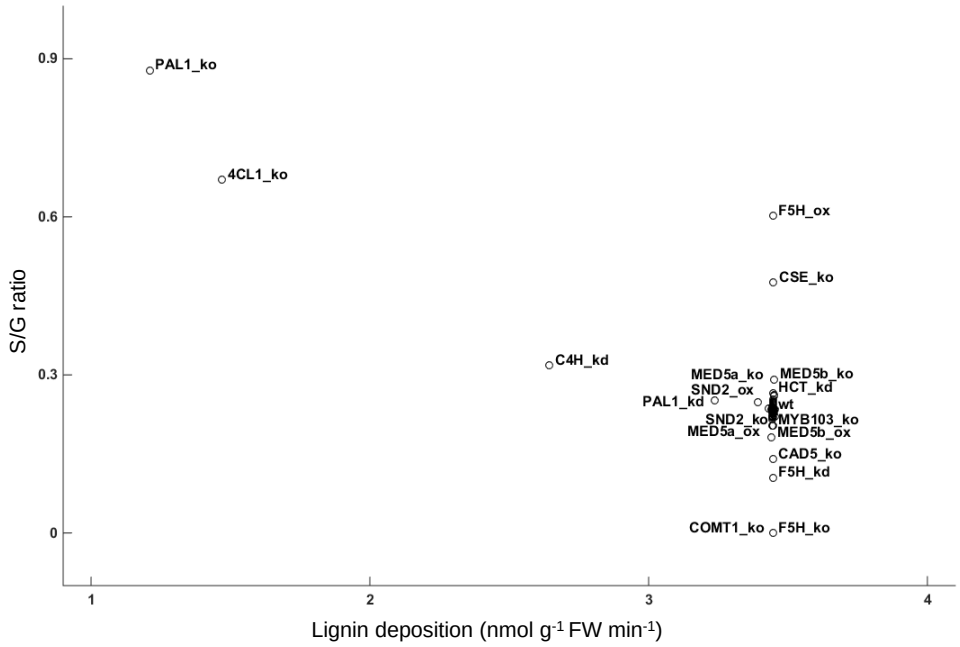

Fig. 4.7. Lignin phenotypes of various genetic perturbating lines predicted by the hierarchical model. The hierarchical model was applied to simulate steady states for various genetic perturbations by multiplying the synthesis of each gene with a factor of zero for knockout (ko), 0.5 for knockdown (kd), and 2 for overexpression (ox), individually. Total lignin deposition rate and the S/G ratio were calculated for each perturbation case after the steady state was reached. Most points were overlapped with wild type condition. Points separated from the point of wild type were labeled with the corresponding genetic perturbations.

states of gene expression were obtained for all cases except for *SND1*-overexpression. A further investigation identified that self-activation of SND1 in the model led to an infinite accumulation of *SND1*, once its synthesis passed a threshold point by overexpression. While self-activation of SND1 was supported by the direct binding of SND1 protein to its own promoter region [52], an unknown mechanism likely exists in plants to prevent its hyperaccumulation to ensure the system robustness. Except for *SND1*-overexpression, each gene expression output from the GRN was then fed into the kinetic model to predict the new metabolic steady state in response to the

corresponding genetic perturbation. Total lignin deposition rate and the S/G ratio were extracted from the model once the steady state was achieved (Fig. 4.7). The highest S/G ratio was reached with *PAL1* knockout, concomitant with around 65% reduction in lignin content, which are qualitatively consistent with the previous findings in *pal1* genotype [35]. Increased S/G ratios were also observed for *4cl1, ref3, cse, F5H*-overexpression lines [18, 21, 31, 209], which were captured by the model as well (Fig. 4.7). The model predicted moderate changes in S/G ratios and lignin content by perturbing TFs in general. The expression of *MED5a/b* was predicted to be inversely proportional to the S/G ratio (Fig. 4.7). However, [22] showed a decreasing trend in S/G ratio in *med5a/b* genotype. On the other hand, the model predicted a correct trend of decreased S/G ratio when *MYB103* is knocked out (Fig. 4.7). Decreased S/G ratios were reported in *myb103* genotypes from [226], which is likely due to the observed dependence of F5H expression on MYB103 [226]. *MYB103* was shown to be upregulated in *SND2*-overexpressing plants [227], which indicates SND2 can also have positive effects on *F5H* expression indirectly. Consistently, the model predicted positive correlation between S/G ratios and *SND2* expression (Fig. 4.7). These observations suggest a hierarchical regulation of SND2 and MYB103 over *F5H*, which is a key metabolic step controlling S/G ratios in Arabidopsis. Overall, these results demonstrate the potential of the hierarchical model in predicting lignin phenotypes for different genetic perturbations, as well as facilitating the global understanding of the lignin biosynthesis.

4.4 Discussion

Systems biology approaches of lignin biosynthesis have been taken for Poplar [51, 92, 174], alfalfa [93], switchgrass [94, 228], etc. for their economic values. On the other hand, [51] presented a list of identified TFs in lignin biosynthesis in various plant species, with the largest number of TFs discovered and characterized in Arabidopsis thaliana, significantly more than any other species. Therefore, although genetic studies

for lignin biosynthesis in Arabidopsis may not be directly transferrable to industrial crops, a more complete view of lignin biosynthesis can be established by integrating the knowledge learned from the previous efforts. With the belief that such systematic understanding will provide a rational guidance for genetic manipulation of other plants, we collected all reported transcriptional/post-transcriptional regulatory interactions in Arabidopsis, and an evidence-derived GRN was constructed based on literature evidence types (Supplementary Fig. C.2). [51] have proposed a model with hierarchical structure to describe the regulation over lignin biosynthesis, which is driven with a top NAC master switch and followed by a series of second or third level master switches such as MYB factors. TFs were described to function in this hierarchical way to regulate lignin biosynthetic genes. Similar feedforward structures have been proposed in the GRNs from [140] as well. However, the literature has examples with a much more complex process for lignin biosynthesis regulation in Arabidopsis. Feedback controls to SND1 from MYB4, MYB7 and MYB32 were reported in [52], with these MYB factors being the downstream targets of other MYB factors such as MYB46 [53, 54] and MYB83 [54]. MYB46 and MYB83 ranked in the top ten TFs with highest number of interactions in the network (Table 4.1). This means *SND1* expression could be indirectly influenced by the combined effects of all these interactions, if most of interactions identified previously are of physiological relevance. Evidence also exists for self-activation for SND1 [52] and self-inhibition for MYB4 and MYB7 [53], meanwhile Mediator complex shows a distal connection with SND1 through a series of MYB factors. All these complex interactions were preserved in the parameterized GRN (Fig. 4.2b and Fig. 4.3), suggesting the gene regulation over lignin biosynthesis proposed by [51] is likely an oversimplified model.

The GRN parameterized with RNA-seq datasets in the current study suggested the Mediator complex is regulating lignin biosynthetic genes by mechanisms other than via MYB factors. Current GRN is not able to propose hypotheses for the possible mechanisms, therefore more studies are needed to investigate the regulatory mechanisms that control the expression of lignin biosynthetic genes. Additionally,

further studies on the regulation of 4CL1 might be warranted, since it ranked the first in the top ten structural genes with highest number of interactions in the network (Table 4.2), suggesting it to be an important gene. Yet direct bindings of TFs were only established for three cases (Supplementary Table C.3), far less than the average evidence ratio of direct binding in the network (10.3% versus 23.6%).

Kinetic modeling offers a mechanistic framework to integrate information from different biological components such as metabolites and enzymes to the emergent properties of the system. It has been shown that given the correctness of individual enzyme kinetics, the system developed *in silico* using kinetic modeling can reproduce the cellular dynamics accurately [91]. While the information of kinetic parameters for enzymes in lignin biosynthesis in Arabidopsis are incomplete, we implemented an optimization approach to estimate all unknown factors simultaneously given data from transcriptome, metabolome, lignin compositions, as well as selected enzyme assays. We compared the performance of two kinetic models, each with V_{max} values tuned from RNA-seq datasets or from the GRN estimation. It should be noted that since our GRN lacks the potential signaling mechanisms from metabolism to transcriptional machinery, TFs expression in our GRN was not affected by the genetic perturbations to the downstream structural genes. Yet such feedback influences were frequently observed for different genetic lines such as in *pal1 pal2* plants [35]. Therefore the estimated gene expression from GRN is only an approximation of the measurements, and a worse performance in data fitting is expected for the model using GRN inputs, if transcriptional regulation plays significant roles in lignin biosynthesis in response to genetic perturbations. While the model with GRN inputs had a worse performance in *cse-2* simulation, it has similar capability in reproducing experimental observations as the model with RNA-seq inputs for other simulation. The model with GRN inputs even had a slightly better match with data from *med5a/b* plants (Fig. 4.4). It is possible this is due to the insufficient training of models with limited size of datasets. An alternative explanation might be that while the expression of genes is sensitive to local perturbations, some post-transcriptional mechanisms exist to partially control

the whole metabolism closer to the wild type condition. Furthermore, understanding of post-transcriptional regulation is likely important and missing in the current model. Information on this could be obtained with proteomic data.

The hierarchical model combining the GRN and the kinetic model reproduced the lignin phenotypes in different genotypes reasonably well (Fig. 4.4). And it also incorporated the enzyme information for PAL, 4CL, HCT and CCR correctly (Fig. 4.5). However, a global underestimation for metabolite concentrations was observed from the model outputs (Supplementary Fig. C.3). Initially the objective function was designed to put less weights for metabolome data, since they are collected from the whole basal stem tissue, thus containing information other than metabolite pools in lignifying cell cytoplasm. Yet such large deviations for so many metabolites were still unexpected. One can argue that shikimate is also present in plastids where it is produced [1]; and *p*-coumarate, caffeate and ferulate have been found to be majorly sequestered in vacuole in soybeans [167]; sinapate is likely accumulated mainly in epidermal cells to generate sinapate esters for UV protection [24, 207–209, 224]; aldehydes/alcohols also likely exist extracellularly, since they are involved in the lignin polymerization. However, these facts fail to explain the underestimation of CoA esters in the model (Supplementary Fig. C.3). On the other hand, the high amounts of CoA esters are indeed contradictory with the high CCR activity (Fig. 4.5a) and high binding affinities of CCR enzymes (Supplementary Table C.1). Note that the highest K_m value of CCR is for caffeoyl CoA, which is reported to be 12.5 μM [36]. Even by an extreme estimation of cytosol volume in basal stem tissue, the highest value of binding affinity of caffeoyl CoA can be no larger than 12.5 nmol g FW^{-1} (assuming 100% of cytosol volume, making the unit conversion factor to be 1 mL g FW^{-1}). The total pool of CoA esters was 9.6 nmol g FW^{-1} (Supplementary Fig. C.3) and CCR maximal capacity was 42.6 nmol g FW^{-1} min^{-1} in wild type Arabidopsis (Fig. 4.5a). Also note that while CCR V_{max} was measured only with *p*-coumaroyl CoA as the substrate, this measurement is in fact a lower bound since *p*-coumaroyl CoA has the slowest k_{cat} among the various CoA esters [36, 37]. Calculating with typical Michaelis-

Menten kinetics, the fluxes through CCR steps should be no smaller than 19 nmol g FW^{-1} min^{-1} given all above information, which is almost six fold of measured lignin deposition rate in wild type Arabidopsis (3.44 nmol g FW^{-1} min^{-1}, Supplementary Fig. C.1). This simple calculation clearly suggests inconsistencies between different type of measurements, and current model was only able to fit lignin fluxes and maximal capacities by sacrificing the performance in reproducing metabolite concentrations. More advanced techniques providing more reliable measurements or finer subcellular resolution are expected to overcome the challenges faced with the current model. Meanwhile it is interesting to explore a more flexible modeling framework to be relatively insensitive to these inconsistencies. Such a framework might be achievable with some more carefully designed weighting mechanisms, which is rudimentarily explored in the current study.

The hierarchical model was able to generate flux maps under various genotypes (Fig. 4.6, Supplementary Fig. C.4 – C.9). The model successfully captured increased S/G ratios in several genotypes [21, 31, 35, 209], without the extra assumptions such as metabolic channeling as taken in some other modeling approaches for lignin biosynthesis [93–95]. Flux towards the S lignin synthesis was robust against genetic perturbations in the upstream pathway, since F5H was functioning at almost its V_{max} in wild type simulation (Fig. 4.5c, 4.6). It is interesting to observe that F5H catalyzed coniferyl alcohol consumption is the only enzyme limiting step in wild type simulation, so that S lignin production is tightly controlled by F5H expression. This tight control was confirmed with the simulation of *F5H*-overexpression and *fah1* (Supplementary Fig. C.8, C.9), with nearly 100% and 0% of S monomer in the final lignin polymer, respectively. In addition, the hierarchical model quantitatively evaluated the contributions from CSE and HCT branches for the synthesis of G and S lignin (Fig. 4.6). 78% of carbon flux from the first HCT step was going through CSE branch in wild type simulation (Fig. 4.6). The high abundances of 4CL enzymes ensured the efficient utilization of caffeate to generate downstream products (Fig. 4.5a). The presence of CSE branch reduced the burden of HCT in the model, since *p*-coumaroyl CoA and

caffeoyl shikimate compete for the HCT catalytic sites. The competition made HCT itself unable to efficiently direct most carbon flux to G and S lignin synthesis, since without CSE, 46% of flux entering lignin biosynthesis ended up as H lignin in the model (Supplementary Fig. C.6). This is in agreement with higher H lignin content in cse plants [21]. Intriguingly, in *pal1 pal2* and *4cl1* simulation, flux through the second HCT step was much less affected in response to the reduction of upstream flux, while flux through CSE was largely reduced (Supplementary Fig. C.4, C.5). This was mainly due the reduction of *p*-coumaroyl CoA led to less competition between *p*-coumaroyl CoA and caffeoyl shikimate over HCT catalytic sites, which increased the relative contribution from second HCT step for G and S lignin synthesis (Supplementary Fig. C.4, C.5). In all simulated genotypes, ALDH steps did not carry any significant fluxes. While ALDH has been found to be important for acid synthesis [229], current modeling results suggest it might play important roles in other cell types such as epidermis.

Isoforms are known to be present for PAL, 4CL and CAD steps for lignin biosynthesis in Arabidopsis [23, 31, 35]. The hierarchical model predicted the majority of fluxes going through one dominant isoform for each step, namely, PAL1, 4CL1 and CAD5. Compensating roles of other isoforms were only observed when the dominant isoform was deleted (Supplementary Fig. C.4, C.5). Shifted flux distribution among acids were observed in *4cl1* simulation (Supplementary Fig. C.5). 4CL catalyzed caffeate consumption was responsible for 70% of synthetic flux of G and S lignin in wild type simulation (Fig. 4.6), while this reduced to only 2% in 4cl1 simulation (Supplementary Fig. C.5). In *4cl1* simulation, ferulate was used instead to contribute to 39% of flux to G and S lignin synthesis (Supplementary Fig. C.5). Examinations over acid concentrations in the model found that all acids were accumulated in *4cl1* simulation, however, *p*-coumarate and ferulate accumulated to 1877 and 252 fold of wild type concentrations, while caffeate only accumulated to 3 fold of wild type case. These huge differences led to the almost full occupation of 4CL catalytic sites with *p*-coumarate and ferulate, thus shifting the flux distribution of acid consumptions in *4cl1* simulation.

For CAD steps, multiple CAD isoforms are found in Arabidopsis in addition to major isoforms CAD4 and CAD5 [33]. The current model only incorporated CAD3 as a third isoform, yet it did not carry any significant fluxes in all simulated cases due to its extremely low expression. *cadC cadD* plants in [32] still deposited G and S lignin, suggesting the important compensating roles of other CAD isoforms. Therefore, incorporations of other CAD isoforms might be necessary to accurately capture the observations in *cadC cadD* plants.

The integration between the GRN and the kinetic model for lignin biosynthesis has great potential in predicting the lignin phenotypes by simulating the cellular metabolism of various genotypes. By simulating the effects of knockout, knockdown and overexpression for each gene in the network, the hierarchical model made predictions for the total lignin contents and S/G ratios for each genotype (Fig. 4.7). Most of the findings were consistent with previous reports, supporting the reliability of the model to reproduce *in vivo* situations. Overall the model predicted that higher S/G ratios can be achieved by either removing upstream enzymes to reduce G lignin synthesis, or overexpressing F5H enzyme to direct more flux to S lignin. Overexpression of *SND2* or knockout of *MYB103* was predicted to have effects on S/G ratio, which is due to the activation of *F5H* expression by their presence in the model. While the effect of *MYB103* knockout has been validated in previous literature [226], it would be interesting to see if *SND2* overexpression also led to higher S/G ratios in Arabidopsis. In general, it seems that perturbing structural genes has more significant effects on lignin phenotypes, however, due to the dynamic and highly-connected nature of the hierarchical model, nonlinear outcomes might exist for combinatory genetic perturbations. This is worth exploring in future studies, and the current model is a great starting point to explore such possibilities. Of course, some improvements can be made for the model presented in the current study, such as resolving the artifacts of complete inhibition of PAL2 by KFBs in wild type, infinite loops of self-activation and accumulation of *SND1* in *SND1*-overexpression lines. These limitations could be solved by having more mechanistic equations for those regulatory processes.

In conclusion, the hierarchical modeling approach by integrating a large scale GRN and a kinetic model offers great opportunities to study the dynamic responses of the metabolism to the transcriptional regulatory machinery and various genetic perturbations. Most of the observed phenotypes in previous reports for lignin biosynthesis in Arabidopsis can be captured by the current model, with explanations provided by the interactions between transcriptional regulation, isoform and substrate competitions. These *in silico* results generate new hypotheses to facilitate a global understanding of the whole pathway, and the model also acts as an efficient platform to systematically predict desired genotypes by scanning possible genetic strategies for all gene targets related to the pathway.

5. SUMMARY

Mathematical modeling of cellular process is an important tool in systems biology, which helps to unravel the dynamic interactions between individual biological components. This is usually achieved by developing models with limited but plausible assumptions, parameterizing the model with data, and then testing the model by making predictions to be verified with experimental measurements. Various analysis can then applied with this *in silico* model for a better understanding of the biological system. For instance, biological properties such as metabolic fluxes, which are usually not able to be directly measured with current techniques, can be estimated with solving a kinetic model. Competing hypotheses of the possible regulatory mechanisms occurring in cells can be tested by mechanistically integrating them into separate kinetic models, and comparison of these candidate model capabilities in reproducing observations can help to reject incorrect hypotheses. Moreover, a dynamic model of complex systems can be a valuable platform to iteratively integrating new measurements and generating testable interpretations/hypotheses over the system, offering great opportunities to identify knowledge gaps, as well as promising directions for further exploration. In the current study, depending on the amount/types of available measurements and questions to answer, various mathematical modeling strategies are taken to simulate metabolism in plants.

Metabolic flux analysis was taken to quantitatively evaluate the contributions of two alternative L-phenylalanine (Phe) biosynthetic routes in Petunia. Due to the limited kinetic knowledge of the enzymes in the pathway, simplified metabolic models were constructed. The models were able to calculate the steady state flow of carbon, metabolic fluxes, which were estimated from isotopic labeled substrate feeding data using mass and isotope balances. We showed that fluxes can be reliably estimated with this metabolic model, and the model helped to reveal the importance of the cytosolic

phenylpyruvate pathway for Phe biosynthesis in Petunia, as well as the potential interactions between two alternative routes for Phe homeostasis. Furthermore, similar metabolic flux analysis in *pal1 pal2* Petunia has determined a significant sequestering pool of Phe when its consumption was impaired. Hyperaccumulation of cytosolic Phe also shuttled down the shikimate pathway in plastids, as suggested by the model. These simple models helped to extract flux information from small datasets, and together with experimental efforts, a complex Phe biosynthetic network in plants is being uncovered.

One limitation of metabolic flux analysis is being unable to simulate dynamic situations. Kinetic modeling is a superior strategy for such cases. This modeling strategy simulates cellular metabolism by representing the kinetics of each enzyme as a differential equation. Interactions between enzymes and pathway intermediates can be mechanistically simulated with this type of model. To study the dynamic responses of the general phenylpropanoid metabolism in Arabidopsis over different metabolic perturbations, we formulated and parameterized a kinetic model to investigate the key mechanisms contributing to the observed metabolic profiles under different isotopically labeled Phe treatments. With an information theoretic approach, a kinetic model with vacuole storage of Phe and *p*-coumarate was shown to have best match with feeding datasets. Plastidial Phe export was predicted to be the major metabolic step to control the flux entering the phenylpropanoid metabolism. Moreover, feedback inhibition from cytosolic Phe toward the plastidial transporter in the model provided one possible way for plastids to react to downstream Phe requirements. *In silico* parameterization of these metabolic steps paved the way for applying similar model developing approach to the whole lignin biosynthesis.

Large scale models have tremendous advantages in capturing complex biological phenomena for a wide range of environmental and genetic conditions. This is because a model with limited coverage needs to have more assumptions to deal with the interactions between modeled players and components outside of modeling scope. These assumptions can be easily invalid if those interactions change dramatically

and bring unexpected influences into the model system. While a large scale model is able to simulate most relevant interactions to cover the complex biological network, sufficient model training is an enormous challenge limiting its success. In the current study, we aimed to incorporate transcriptional control of lignification in Arabidopsis by constructing a hierarchical model combining a gene regulatory network and the kinetic model of lignin biosynthesis. To overcome the curse of dimensionality in large scale modeling, attempts have been made to take full use of established knowledge for the pathway. A gene regulatory network was first constructed from a candidate interaction sets with reported literature evidence. Next, a kinetic model for the entire lignin biosynthetic pathway downstream of Phe was formulated using *in vitro* information and by estimating kinetic parameters with missing reference values. In this way, we managed to develop a large scale model with data from 13 different genotypes in Arabidopsis. The model contains hundreds of interactions and reactions, and it successfully reproduced observed phenotypes in various genetic perturbation lines. Detailed explanations are provided to uncover the causes for those observations, which can be further explored either with new experiments or testing the model with new datasets. The model was also applied to thoroughly examine for promising genotypes for desired phenotypes resulting from alterations of single gene expression, which opened more possibilities for metabolic engineering.

Overall, mathematical modeling is a powerful approach to find a multifaceted in silico framework that integrates information from data, even if they are from different types of measurements. The flexibility of the modeling approaches makes it suitable to meet different needs to interpret data for different questions and applications. Here we have shown that mathematical modeling can advance the understanding of plant metabolism by learning from data to propose casual relationships between different metabolic factors. Modeling applications in Phe and lignin biosynthetic networks aided in discoveries of key mechanisms in cellular metabolism, and extension of these modeling efforts are expected to continually contribute to answer different research questions in plants.

REFERENCES

REFERENCES

[1] H. Maeda and N. Dudareva, "The shikimate pathway and aromatic amino acid biosynthesis in plants," *Annual Review of Plant Biology*, vol. 63, pp. 73–105, 2012.

[2] E. Haslam, *Shikimic acid: metabolism and metabolites*. John Wiley & Sons Inc, 1993.

[3] D. L. Siehl, "The biosynthesis of tryptophan, tyrosine. and phenylalanine from cborismate," in *Plant Amino Acids*. CRC Press, 1998, pp. 185–218.

[4] H. Maeda, A. K. Shasany, J. Schnepp, I. Orlova, G. Taguchi, B. R. Cooper, D. Rhodes, E. Pichersky, and N. Dudareva, "RNAi suppression of arogenate dehydratase 1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals," *The Plant Cell*, vol. 22, no. 3, pp. 832–849, 2010.

[5] V. Dal Cin, D. M. Tieman, T. Tohge, R. McQuinn, R. C. de Vos, S. Osorio, E. A. Schmelz, M. G. Taylor, M. T. Smits-Kroon, R. C. Schuurink *et al.*, "Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit," *The Plant Cell*, vol. 23, no. 7, pp. 2738–2753, 2011.

[6] O. R. Corea, C. Ki, C. L. Cardenas, S.-J. Kim, S. E. Brewer, A. M. Patten, L. B. Davin, and N. G. Lewis, "Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins," *Journal of Biological Chemistry*, vol. 287, no. 14, pp. 11 446–11 459, 2012.

[7] H. Yoo, J. R. Widhalm, Y. Qian, H. Maeda, B. R. Cooper, A. S. Jannasch, I. Gonda, E. Lewinsohn, D. Rhodes, and N. Dudareva, "An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine: phenylpyruvate aminotransferase," *Nature Communications*, vol. 4, p. 2833, 2013.

[8] G. W. Kuroki and E. E. Conn, "Purification and characterization of an inducible aromatic amino acid-sensitive form of chorismate mutase from Solanum tuberosum L. tubers," *Archives of Biochemistry and Biophysics*, vol. 260, no. 2, pp. 616 – 621, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0003986188904894

[9] M. Benesova and R. Bode, "Chorismate mutase isoforms from seeds and seedlings of Papaver somniferum," *Phytochemistry*, vol. 31, no. 9, pp. 2983 – 2987, 1992, the International Journal of Plant Biochemistry. [Online]. Available: http://www.sciencedirect.com/science/article/pii/003194229283431W

[10] D. L. Siehl and E. E. Conn, "Kinetic and regulatory properties of arogenate dehydratase in seedlings of Sorghum bicolor (L.) moench," *Archives of Biochemistry and Biophysics*, vol. 260, no. 2, pp. 822 – 829, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0003986188905139

[11] T. Yamada, F. Matsuda, K. Kasai, S. Fukuoka, K. Kitamura, Y. Tozawa, H. Miyagawa, and K. Wakasa, "Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan," *The Plant Cell*, vol. 20, no. 5, pp. 1316–1329, 2008. [Online]. Available: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.057455

[12] A. Graziana and A. M. Boudet, "3-Deoxy-D-arabino heptulosonate 7-phosphate synthase from Zea mays: General properties and regulation by tryptophan," *Plant and Cell Physiology*, vol. 21, no. 5, pp. 793–802, 08 1980. [Online]. Available: https://doi.org/10.1093/oxfordjournals.pcp.a076054

[13] M. Reinink and A. Borstlap, "3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase from pea leaves: Inhibition by L-tyrosine," *Plant Science Letters*, vol. 26, no. 2, pp. 167 – 171, 1982. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304421182900888

[14] J. C. Verdonk, M. A. Haring, A. J. van Tunen, and R. C. Schuurink, "ODORANT1 regulates fragrance biosynthesis in petunia flowers," *The Plant Cell*, vol. 17, no. 5, pp. 1612–1624, 2005. [Online]. Available: http://www.plantcell.org/content/17/5/1612

[15] K. Radotić, M. Mićić, and M. Jeremić, "New insights into the structural organization of the plant polymer lignin," *Annals of the New York Academy of Sciences*, vol. 1048, no. 1, pp. 215–229, 2005. [Online]. Available: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1196/annals.1342.020

[16] M. Li, Y. Pu, and A. J. Ragauskas, "Current understanding of the correlation of lignin structure with biomass recalcitrance," *Frontiers in Chemistry*, vol. 4, p. 45, 2016.

[17] P. Wang, N. Dudareva, J. A. Morgan, and C. Chapple, "Genetic manipulation of lignocellulosic biomass for bioenergy," *Current Opinion in Chemical Biology*, vol. 29, pp. 32–39, 2015.

[18] K. Meyer, A. M. Shirley, J. C. Cusumano, D. A. Bell-Lelong, and C. Chapple, "Lignin monomer composition is determined by the expression of a cytochrome p450-dependent monooxygenase in Arabidopsis," *Proceedings of the National Academy of Sciences*, vol. 95, no. 12, pp. 6619–6623, 1998.

[19] F. Chen and R. A. Dixon, "Lignin modification improves fermentable sugar yields for biofuel production," *Nature Biotechnology*, vol. 25, no. 7, p. 759, 2007.

[20] R. Van Acker, R. Vanholme, V. Storme, J. C. Mortimer, P. Dupree, and W. Boerjan, "Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in Arabidopsis thaliana," *Biotechnology for Biofuels*, vol. 6, no. 1, p. 46, 2013.

[21] R. Vanholme, I. Cesarino, K. Rataj, Y. Xiao, L. Sundin, G. Goeminne, H. Kim, J. Cross, K. Morreel, P. Araujo *et al.*, "Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis," *Science*, vol. 341, no. 6150, pp. 1103–1106, 2013.

[22] N. D. Bonawitz, J. Im Kim, Y. Tobimatsu, P. N. Ciesielski, N. A. Anderson, E. Ximenes, J. Maeda, J. Ralph, B. S. Donohoe, M. Ladisch *et al.*, "Disruption of Mediator rescues the stunted growth of a lignin-deficient arabidopsis mutant," *Nature*, vol. 509, no. 7500, p. 376, 2014.

[23] N. A. Anderson, Y. Tobimatsu, P. N. Ciesielski, E. Ximenes, J. Ralph, B. S. Donohoe, M. Ladisch, and C. Chapple, "Manipulation of guaiacyl and syringyl monomer biosynthesis in an Arabidopsis cinnamyl alcohol dehydrogenase mutant results in atypical lignin biosynthesis and modified cell wall structure," *The Plant Cell*, vol. 27, no. 8, pp. 2195–2209, 2015.

[24] R. Franke, M. R. Hemm, J. W. Denault, M. O. Ruegger, J. M. Humphreys, and C. Chapple, "Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis," *The Plant Journal*, vol. 30, no. 1, pp. 47–59, 2002.

[25] L. Hoffmann, S. Besseau, P. Geoffroy, C. Ritzenthaler, D. Meyer, C. Lapierre, B. Pollet, and M. Legrand, "Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis," *The Plant Cell*, vol. 16, no. 6, pp. 1446–1465, 2004.

[26] N. Abdulrazzak, B. Pollet, J. Ehlting, K. Larsen, C. Asnaghi, S. Ronseau, C. Proux, M. Erhardt, V. Seltzer, J.-P. Renou *et al.*, "A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth," *Plant Physiology*, vol. 140, no. 1, pp. 30–48, 2006.

[27] H. D. Coleman, A. L. Samuels, R. D. Guy, and S. D. Mansfield, "Perturbed lignification impacts tree growth in hybrid poplara function of sink strength, vascular integrity, and photosynthetic assimilation," *Plant Physiology*, vol. 148, no. 3, pp. 1229–1237, 2008.

[28] N. D. Bonawitz and C. Chapple, "The genetics of lignin biosynthesis: Connecting genotype to phenotype," *Annual Review of Genetics*, vol. 44, no. 1, pp. 337–363, 2010. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev-genet-102209-163508

[29] J. Ehlting, D. Büttner, Q. Wang, C. J. Douglas, I. E. Somssich, and E. Kombrink, "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms," *Plant Journal*, vol. 19, no. 1, pp. 9–20, 1999.

[30] B. Hamberger and K. Hahlbrock, "The 4-coumarate: CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes," *Proceedings of the National Academy of Sciences*, vol. 101, no. 7, pp. 2209–2214, 2004.

[31] Y. Li, J. I. Kim, L. Pysh, and C. Chapple, "Four isoforms of Arabidopsis thaliana 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism," *Plant Physiology*, vol. 4, no. 765, pp. 2409–2421, 2015. [Online]. Available: http://www.plantphysiol.org/lookup/doi/10.1104/pp.15.00838

[32] R. Sibout, A. Eudes, B. Pollet, T. Goujon, I. Mila, F. Granier, A. Séguin, C. Lapierre, and L. Jouanin, "Expression pattern of two paralogs encoding cinnamyl alcohol dehydrogenases in Arabidopsis. isolation and characterization of the corresponding mutants," *Plant Physiology*, vol. 132, no. 2, pp. 848–860, 2003.

[33] S.-J. Kim, M.-R. Kim, D. L. Bedgar, S. G. Moinuddin, C. L. Cardenas, L. B. Davin, C. Kang, and N. G. Lewis, "Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis," *Proceedings of the National Academy of Sciences*, vol. 101, no. 6, pp. 1455–1460, 2004.

[34] F. C. Cochrane, L. B. Davin, and N. G. Lewis, "The Arabidopsis phenylalanine ammonia lyase gene family: Kinetic characterization of the four PAL isoforms," *Phytochemistry*, vol. 65, no. 11, pp. 1557–1564, 2004.

[35] A. Rohde, K. Morreel, J. Ralph, G. Goeminne, V. Hostyn, R. De Rycke, S. Kushnir, J. Van Doorsselaere, J.-P. Joseleau, M. Vuylsteke, G. Van Driessche, J. Van Beeumen, E. Messens, and W. Boerjan, "Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism," *The Plant Cell*, vol. 16, no. 10, pp. 2749–2771, 2004. [Online]. Available: http://www.plantcell.org/content/16/10/2749

[36] V. Lauvergeat, C. Lacomme, E. Lacombe, E. Lasserre, D. Roby, and J. Grima-Pettenati, "Two cinnamoyl-CoA reductase (CCR) genes from Arabidopsis thaliana are differentially expressed during development and in response to infection with pathogenic bacteria," *Phytochemistry*, vol. 57, no. 7, pp. 1187–1195, 2001.

[37] M. Baltas, C. Lapeyre, F. Bedos-Belval, M. Maturano, P. Saint-Aguet, L. Roussel, H. Duran, and J. Grima-Pettenati, "Kinetic and inhibition studies of cinnamoyl-CoA reductase 1 from Arabidopsis thaliana," *Plant Physiology and Biochemistry*, vol. 43, no. 8, pp. 746–753, 2005.

[38] N. A. Anderson, N. D. Bonawitz, K. Nyffeler, and C. Chapple, "Loss of ferulate 5-hydroxylase leads to mediator-dependent inhibition of soluble phenylpropanoid biosynthesis in Arabidopsis," *Plant Physiology*, vol. 169, no. 3, pp. 1557–1567, 2015.

[39] L. Hoffmann, S. Maury, F. Martz, P. Geoffroy, and M. Legrand, "Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism," *Journal of Biological Chemistry*, vol. 278, no. 1, pp. 95–103, 2003.

[40] J. M. Humphreys, M. R. Hemm, and C. Chapple, "New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase," *Proceedings of the National Academy of Sciences*, vol. 96, no. 18, pp. 10 045–10 050, 1999.

[41] S. G. Moinuddin, M. Jourdes, D. D. Laskar, C. Ki, C. L. Cardenas, K.-W. Kim, D. Zhang, L. B. Davin, and N. G. Lewis, "Insights into lignin primary structure and deconstruction from Arabidopsis thaliana COMT (caffeic acid O-methyl transferase) mutant Atomt1," *Organic & Biomolecular Chemistry*, vol. 8, no. 17, pp. 3928–3946, 2010.

[42] R. Zhong and Z.-H. Ye, "Complexity of the transcriptional network controlling secondary wall biosynthesis," *Plant Science*, vol. 229, pp. 193–207, 2014.

[43] X. Rao and R. A. Dixon, "Current models for transcriptional regulation of secondary cell wall biosynthesis in grasses," *Frontiers in Plant Science*, vol. 9, p. 399, 2018.

[44] N. Mitsuda, A. Iwase, H. Yamamoto, M. Yoshida, M. Seki, K. Shinozaki, and M. Ohme-Takagi, "NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis," *The Plant Cell*, vol. 19, no. 1, pp. 270–280, 2007.

[45] K. Ohashi-Ito, Y. Oda, and H. Fukuda, "Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation," *The Plant Cell*, vol. 22, no. 10, pp. 3461–3473, 2010.

[46] M. Yamaguchi, N. Mitsuda, M. Ohtani, M. Ohme-Takagi, K. Kato, and T. Demura, "VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation," *The Plant Journal*, vol. 66, no. 4, pp. 579–590, 2011.

[47] Y.-C. Lin, W. Li, Y.-H. Sun, S. Kumari, H. Wei, Q. Li, S. Tunlaya-Anukit, R. R. Sederoff, and V. L. Chiang, "SND1 transcription factor–directed quantitative functional hierarchical genetic regulatory network in wood formation in Populus trichocarpa," *The Plant cell*, vol. 25, no. 11, pp. 4324–4341, 2013.

[48] J. Zhou, R. Zhong, and Z.-H. Ye, "Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels," *PloS ONE*, vol. 9, no. 8, p. e105726, 2014.

[49] R. Zhong and Z.-H. Ye, "The Arabidopsis NAC transcription factor NST2 functions together with SND1 and NST1 to regulate secondary wall biosynthesis in fibers of inflorescence stems," *Plant Signaling & Behavior*, vol. 10, no. 2, p. e989746, 2015.

[50] W. Xiao, Y. Yang, and J. Yu, "ZmNST3 and ZmNST4 are master switches for secondary wall deposition in maize (Zea mays L.)," *Plant Science*, vol. 266, pp. 83–94, 2018.

[51] H. Chen, J. P. Wang, H. Liu, H. Li, Y.-C. J. Lin, S. Rui, C. Yang, J. Gao, C. Zhou, Q. Li *et al.*, "Hierarchical transcription-factor and chromatin binding network for wood formation in Populus trichocarpa," *The Plant Cell*, pp. tpc–00 620, 2019.

[52] H. Wang, Q. Zhao, F. Chen, M. Wang, and R. A. Dixon, "NAC domain function and transcriptional control of a secondary cell wall master switch," *The Plant Journal*, vol. 68, no. 6, pp. 1104–1114, 2011.

[53] J.-H. Ko, W.-C. Kim, and K.-H. Han, "Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in Arabidopsis," *The Plant Journal*, vol. 60, no. 4, pp. 649–665, 2009.

[54] R. Zhong and Z.-H. Ye, "MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes," *Plant and Cell Physiology*, vol. 53, no. 2, pp. 368–380, 2011.

[55] R. Zhong, E. A. Richardson, and Z.-H. Ye, "The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis," *The Plant Cell*, vol. 19, no. 9, pp. 2776–2792, 2007.

[56] R. Zhong, C. Lee, J. Zhou, R. L. McCarthy, and Z.-H. Ye, "A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis," *The Plant Cell*, vol. 20, no. 10, pp. 2763–2782, 2008.

[57] S. Fornalé, E. Lopez, J. E. Salazar-Henao, P. Fernández-Nohales, J. Rigau, and D. Caparros-Ruiz, "AtMYB7, a new player in the regulation of UV-sunscreens in Arabidopsis thaliana," *Plant and Cell Physiology*, vol. 55, no. 3, pp. 507–516, 2014.

[58] W. L. Dolan, B. P. Dilkes, J. M. Stout, N. D. Bonawitz, and C. Chapple, "Mediator complex subunits MED2, MED5, MED16, and MED23 genetically interact in the regulation of phenylpropanoid biosynthesis," *The Plant Cell*, vol. 29, no. 12, pp. 3269–3285, 2017.

[59] X. Zhang, M. Gou, and C.-J. Liu, "Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase," *The Plant Cell*, vol. 25, no. 12, pp. 4994–5010, 2013. [Online]. Available: http://www.plantcell.org/cgi/doi/10.1105/tpc.113.119644

[60] X. Zhang, M. Gou, C. Guo, H. Yang, and C.-J. Liu, "Down-regulation of Kelch domain-containing F-box protein in Arabidopsis enhances the production of (poly)phenols and tolerance to ultraviolet radiation," *Plant Physiology*, vol. 167, no. 2, pp. 337–350, 2015. [Online]. Available: http://www.plantphysiol.org/lookup/doi/10.1104/pp.114.249136

[61] X. Zhang and C. J. Liu, "Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids," *Molecular Plant*, vol. 8, no. 1, pp. 17–27, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.molp.2014.11.001

[62] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, p. 57, 2009.

[63] L. Gao, J. Wang, H. Ge, L. Fang, Y. Zhang, X. Huang, and Y. Wang, "Toward the complete proteome of Synechocystis sp. PCC 6803," *Photosynthesis Research*, vol. 126, no. 2-3, pp. 203–219, 2015.

[64] C. Ramus, A. Hovasse, M. Marcellin, A.-M. Hesse, E. Mouton-Barbosa, D. Bouyssié, S. Vaca, C. Carapito, K. Chaoui, C. Bruley *et al.*, "Benchmarking quantitative label-free LC–MS data processing workflows using a complex spiked proteomic standard dataset," *Journal of Proteomics*, vol. 132, pp. 51–62, 2016.

[65] R. Jaini, P. Wang, N. Dudareva, C. Chapple, and J. A. Morgan, "Targeted metabolomics of the phenylpropanoid pathway in Arabidopsis thaliana using reversed phase liquid chromatography coupled with tandem mass spectrometry," *Phytochemical Analysis*, vol. 28, no. 4, pp. 267–276, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pca.2672

[66] W. Wiechert, "^{13}C metabolic flux analysis," *Metabolic Engineering*, vol. 3, no. 3, pp. 195–206, 2001.

[67] M. R. Antoniewicz, "Methods and advances in metabolic flux analysis: a mini-review," *Journal of Industrial Microbiology & Biotechnology*, vol. 42, no. 3, pp. 317–325, 2015.

[68] W. Wiechert and A. De Graaf, "In vivo stationary flux analysis by ^{13}C labeling experiments," in *Metabolic Engineering*. Springer, 1996, pp. 109–154.

[69] J. D. Young, A. A. Shastri, G. Stephanopoulos, and J. A. Morgan, "Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis," *Metabolic Engineering*, vol. 13, no. 6, pp. 656–665, 2011.

[70] J. D. Young, "INCA: a computational platform for isotopically non-stationary metabolic flux analysis," *Bioinformatics*, vol. 30, no. 9, pp. 1333–1335, 2014.

[71] S. B. Crown, C. P. Long, and M. R. Antoniewicz, "Integrated ^{13}C-metabolic flux analysis of 14 parallel labeling experiments in Escherichia coli," *Metabolic Engineering*, vol. 28, pp. 151–158, 2015.

[72] L. T. Cordova and M. R. Antoniewicz, "^{13}C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing Geobacillus strain LC300," *Metabolic Engineering*, vol. 33, pp. 148–157, 2016.

[73] S. Gopalakrishnan and C. D. Maranas, "^{13}C metabolic flux analysis at a genome-scale," *Metabolic Engineering*, vol. 32, pp. 12–22, 2015.

[74] F. Ma, L. J. Jazmin, J. D. Young, and D. K. Allen, "Isotopically nonstationary ^{13}C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation," *Proceedings of the National Academy of Sciences*, vol. 111, no. 47, pp. 16 967–16 972, 2014.

[75] Y. Hanba, H. Kogami, and I. Terashima, "The effect of growth irradiance on leaf anatomy and photosynthesis in Acer species differing in light demand," *Plant, Cell & Environment*, vol. 25, no. 8, pp. 1021–1030, 2002.

[76] J. Schwender, I. Hebbelmann, N. Heinzel, T. Hildebrandt, A. Rogers, D. Naik, M. Klapperstück, H.-P. Braun, F. Schreiber, P. Denolf *et al.*, "Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture," *Plant Physiology*, vol. 168, no. 3, pp. 828–848, 2015.

[77] X. Zhang, A. Misra, S. Nargund, G. D. Coleman, and G. Sriram, "Concurrent isotope-assisted metabolic flux analysis and transcriptome profiling reveal responses of poplar cells to altered nitrogen and carbon supply," *The Plant Journal*, vol. 93, no. 3, pp. 472–488, 2018.

[78] D. K. Allen, "Quantifying plant phenotypes with isotopic labeling & metabolic flux analysis," *Current Opinion in Biotechnology*, vol. 37, pp. 45–52, 2016.

[79] J. R. Widhalm, M. Gutensohn, H. Yoo, F. Adebesin, Y. Qian, L. Guo, R. Jaini, J. H. Lynch, R. M. McCoy, J. T. Shreve, J. Thimmapuram, D. Rhodes, J. A. Morgan, and N. Dudareva, "Identification of a plastidial phenylalanine exporter that influences flux distribution through the phenylalanine biosynthetic network," *Nature Communications*, vol. 6, p. 8142, 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/ncomms9142

[80] J. H. Lynch, I. Orlova, C. Zhao, L. Guo, R. Jaini, H. Maeda, T. Akhtar, J. Cruz-Lebron, D. Rhodes, J. Morgan *et al.*, "Multifaceted plant responses to circumvent Phe hyperaccumulation by downregulation of flux through the shikimate pathway and by vacuolar Phe sequestration," *Plant Journal*, vol. 92, no. 5, pp. 939–950, 2017.

[81] Y. Qian, J. H. Lynch, L. Guo, D. Rhodes, J. A. Morgan, and N. Dudareva, "Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants," *Nature Communications*, vol. 10, no. 1, p. 15, 2019.

[82] J. A. Morgan and D. Rhodes, "Mathematical modeling of plant metabolic pathways," *Metabolic Engineering*, vol. 4, no. 1, pp. 80–89, 2002.

[83] J. M. Rohwer, "Kinetic modelling of plant metabolic pathways," *Journal of Experimental Botany*, vol. 63, no. 6, pp. 2275–2292, 2012.

[84] L. Menten and M. Michaelis, "Die kinetik der invertinwirkung," *Biochem Z*, vol. 49, no. 333-369, p. 5, 1913.

[85] K. A. Johnson and R. S. Goody, "The original Michaelis constant: translation of the 1913 Michaelis–Menten paper," *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, 2011.

[86] A. V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves," *J. Physiol.*, vol. 40, pp. 4–7, 1910.

[87] J. N. Weiss, "The Hill equation revisited: uses and misuses," *The FASEB Journal*, vol. 11, no. 11, pp. 835–841, 1997.

[88] C. M. Guldberg and P. Waage, "Experiments concerning chemical affinity," *German translation by Abegg in Ostwalds Klassiker der Exacten Wissenshaften*, no. 104, pp. 10–125, 1899.

[89] F. Horn and R. Jackson, "General mass action kinetics," *Archive for Rational Mechanics and Analysis*, vol. 47, no. 2, pp. 81–116, 1972.

[90] E. O. Voit, "Modelling metabolic networks using power-laws and S-systems," *Essays in Biochemistry*, vol. 45, pp. 29–40, 2008.

[91] G. Curien, O. Bastien, M. Robert-Genthon, A. Cornish-Bowden, M. L. Cárdenas, and R. Dumas, "Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters," *Molecular Systems Biology*, vol. 5, no. 271, 2009. [Online]. Available: http://msb.embopress.org/cgi/doi/10.1038/msb.2009.29

[92] Y. Lee and E. O. Voit, "Mathematical modeling of monolignol biosynthesis in Populus xylem," *Mathematical Biosciences*, vol. 228, no. 1, pp. 78 – 89, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0025556410001410

[93] Y. Lee, L. Escamilla-Treviño, R. A. Dixon, and E. O. Voit, "Functional analysis of metabolic channeling and regulation in lignin biosynthesis: A computational approach," *PLoS Computational Biology*, vol. 8, no. 11, 2012.

[94] M. Faraji, L. L. Fonseca, L. Escamilla-Treviño, R. A. Dixon, and E. O. Voit, "Computational inference of the structure and regulation of the lignin pathway in Panicum virgatum," *Biotechnology for Biofuels*, vol. 8, no. 1, p. 151, 2015. [Online]. Available: http://www.biotechnologyforbiofuels.com/content/8/1/151

[95] M. Faraji, L. L. Fonseca, L. Escamilla-Treviño, J. Barros-Rios, N. L. Engle, Z. K. Yang, T. J. Tschaplinski, R. A. Dixon, and E. O. Voit, "A dynamic model of lignin biosynthesis in Brachypodium distachyon," *Biotechnology for Biofuels*, vol. 11, no. 1, p. 253, Sep 2018. [Online]. Available: https://doi.org/10.1186/s13068-018-1241-6

[96] L. Uys, F. C. Botha, J.-H. S. Hofmeyr, and J. M. Rohwer, "Kinetic model of sucrose accumulation in maturing sugarcane culm tissue," *Phytochemistry*, vol. 68, no. 16, pp. 2375 – 2392, 2007, dynamic Metabolic Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031942207002798

[97] X.-G. Zhu, E. de Sturler, and S. P. Long, "Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm," *Plant Physiology*, vol. 145, no. 2, pp. 513–526, 2007. [Online]. Available: http://www.plantphysiol.org/cgi/doi/10.1104/pp.107.103713

[98] Y. Wang, S. P. Long, and X.-G. Zhu, "Elements required for an efficient NADP-malic enzyme type C4 photosynthesis," *Plant Physiology*, vol. 164, no. 4, pp. 2231–2246, 2014. [Online]. Available: http://www.plantphysiol.org/content/164/4/2231

[99] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. Van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. Van Dam, H. V. Westerhoff *et al.*, "Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry," *European Journal of Biochemistry*, vol. 267, no. 17, pp. 5313–5329, 2000.

[100] E. Heinzle, F. Matsuda, H. Miyagawa, K. Wakasa, and T. Nishioka, "Estimation of metabolic fluxes, expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation," *The Plant Journal*, vol. 50, no. 1, pp. 176–187, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2007.03037.x

[101] J. Liu, M. Brazier-Hicks, and R. Edwards, "A kinetic model for the metabolism of the herbicide safener fenclorim in Arabidopsis thaliana," *Biophysical Chemistry*, vol. 143, no. 1, pp. 85 – 94, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301462209000854

[102] K. M. Olsen, R. Slimestad, U. S. Lea, C. Brede, T. LØVDAL, P. Ruoff, M. Verheul, and C. Lillo, "Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies," *Plant, Cell & Environment*, vol. 32, no. 3, pp. 286–299, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2008.01920.x

[103] A. M. Colón, N. Sengupta, D. Rhodes, N. Dudareva, and J. Morgan, "A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of Petunia hybrida," *Plant Journal*, vol. 62, no. 1, pp. 64–76, 2010.

[104] M. Cloutier, M. Perrier, and M. Jolicoeur, "Dynamic flux cartography of hairy roots primary metabolism," *Phytochemistry*, vol. 68, no. 16, pp. 2393 – 2404, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031942207002907

[105] T. Nägele, S. Henkel, I. Hörmiller, T. Sauter, O. Sawodny, M. Ederer, and A. G. Heyer, "Mathematical modeling of the central carbohydrate metabolism in Arabidopsis reveals a substantial regulatory influence of vacuolar invertase on whole plant carbon metabolism," *Plant Physiology*, vol. 153, no. 1, pp. 260–272, 2010. [Online]. Available: http://www.plantphysiol.org/content/153/1/260

[106] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophysical Journal*, vol. 95, no. 12, pp. 5606–5617, 2008.

[107] J. A. Vrugt, "Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation," *Environmental Modelling & Software*, vol. 75, pp. 273–316, 2016.

[108] H. Haario, M. Laine, A. Mira, and E. Saksman, "DRAM: Efficient adaptive MCMC," *Statistics and Computing*, vol. 16, no. 4, pp. 339–354, 2006.

[109] L. Guo, P. Wang, R. Jaini, N. Dudareva, C. Chapple, and J. A. Morgan, "Dynamic modeling of subcellular phenylpropanoid metabolism in Arabidopsis lignifying cells," *Metabolic Engineering*, vol. 49, pp. 36–46, 2018.

[110] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," 1970.

[111] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[112] H. Kacser and J. Burns, "Rate control of biological processes," in *Symp. Soc. Exp. Biol*, vol. 27, 1973, pp. 65–104.

[113] D. A. Fell, "Metabolic control analysis," *Systems Biology*, vol. 13, no. May, pp. 69–80, 2005. [Online]. Available: http://dx.doi.org/10.1007/b137745

[114] R. Rios-Estepa, I. Lange, J. M. Lee, and B. M. Lange, "Mathematical modeling-guided evaluation of biochemical, developmental, environmental, and genotypic determinants of essential oil composition and yield in peppermint leaves," *Plant Physiology*, vol. 152, no. 4, pp. 2105–2119, 2010. [Online]. Available: http://www.plantphysiol.org/content/152/4/2105

[115] B. P. Beauvoit, S. Colombie, A. Monier, M.-H. Andrieu, B. Biais, C. Benard, C. Cheniclet, M. Dieuaide-Noubhani, C. Nazaret, J.-P. Mazat, and Y. Gibon, "Model-assisted analysis of sugar metabolism throughout tomato fruit development reveals enzyme and carrier properties in relation to vacuole expansion," *The Plant Cell*, vol. 26, no. 8, pp. 3224–3242, 2014. [Online]. Available: http://www.plantcell.org/cgi/doi/10.1105/tpc.114.127761

[116] B. Knoke, S. Textor, J. Gershenzon, and S. Schuster, "Mathematical modelling of aliphatic glucosinolate chain length distribution in Arabidopsis thaliana leaves," *Phytochemistry Reviews*, vol. 8, no. 1, p. 39, Aug 2008. [Online]. Available: https://doi.org/10.1007/s11101-008-9107-3

[117] M. Banf and S. Y. Rhee, "Computational inference of gene regulatory networks: approaches, limitations and opportunities," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1860, no. 1, pp. 41–52, 2017.

[118] S. Haque, J. S. Ahmad, N. M. Clark, C. M. Williams, and R. Sozzani, "Computational prediction of gene regulatory networks in plant growth and development," *Current Opinion in Plant Biology*, vol. 47, pp. 96–105, 2019.

[119] G. Sanguinetti, "Gene regulatory network inference: an introductory survey," in *Gene Regulatory Networks*. Springer, 2019, pp. 1–23.

[120] H. Kim, J. K. Lee, and T. Park, "Inference of large-scale gene regulatory networks using regression-based network approach," *Journal of Bioinformatics and Computational Biology*, vol. 7, no. 04, pp. 717–735, 2009.

[121] A.-C. Haury, F. Mordelet, P. Vera-Licona, and J.-P. Vert, "TIGRESS: trustful inference of gene regulation using stability selection," *BMC Systems Biology*, vol. 6, no. 1, p. 145, 2012.

[122] Z. Dong, T. Song, and C. Yuan, "Inference of gene regulatory networks from genetic perturbations with linear regression model," *PloS ONE*, vol. 8, no. 12, p. e83263, 2013.

[123] J. Pirgazi and A. R. Khanteymoori, "A robust gene regulatory network inference method base on Kalman filter and linear regression," *PloS ONE*, vol. 13, no. 7, p. e0200094, 2018.

[124] D. Trejo Banos, A. J. Millar, and G. Sanguinetti, "A Bayesian approach for structure learning in oscillating regulatory networks," *Bioinformatics*, vol. 31, no. 22, pp. 3617–3624, 2015.

[125] M. A. de Luis Balaguer, A. P. Fisher, N. M. Clark, M. G. Fernandez-Espinosa, B. K. Möller, D. Weijers, J. U. Lohmann, C. Williams, O. Lorenzo, and R. Sozzani, "Predicting gene regulatory networks by combining spatial and temporal gene expression data in Arabidopsis root stem cells," *Proceedings of the National Academy of Sciences*, vol. 114, no. 36, pp. E7632–E7640, 2017.

[126] M. N. Zeilinger, E. M. Farré, S. R. Taylor, S. A. Kay, and F. J. Doyle, "A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9," *Molecular Systems Biology*, vol. 2, no. 1, p. 58, 2006.

[127] J. Cao, X. Qi, and H. Zhao, "Modeling gene regulation networks using ordinary differential equations," in *Next Generation Microarray Bioinformatics*. Springer, 2012, pp. 185–197.

[128] F. Dondelinger, D. Husmeier, S. Rogers, and M. Filippone, "ODE parameter inference using adaptive gradient matching with Gaussian processes," in *Artificial Intelligence and Statistics*, 2013, pp. 216–228.

[129] M. Grzegorczyk, A. Aderhold, and D. Husmeier, "Inferring bi-directional interactions between circadian clock genes and metabolism with model ensembles," *Statistical Applications in Genetics and Molecular Biology*, vol. 14, no. 2, pp. 143–167, 2015.

[130] F. L. Valentim, S. van Mourik, D. Posé, M. C. Kim, M. Schmid, R. C. van Ham, M. Busscher, G. F. Sanchez-Perez, J. Molenaar, G. C. Angenent *et al.*, "A quantitative and dynamic model of the Arabidopsis flowering time gene regulatory network," *Plos ONE*, vol. 10, no. 2, p. e0116973, 2015.

[131] K. A. McGoff, X. Guo, A. Deckard, C. M. Kelliher, A. R. Leman, L. J. Francey, J. B. Hogenesch, S. B. Haase, and J. L. Harer, "The local edge machine: inference of dynamic models of gene regulation," *Genome Biology*, vol. 17, no. 1, p. 214, 2016.

[132] F. Mordelet and J.-P. Vert, "SIRENE: supervised inference of regulatory networks," *Bioinformatics*, vol. 24, no. 16, pp. i76–i82, 2008.

[133] A. Greenfield, A. Madar, H. Ostrer, and R. Bonneau, "DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models," *PloS ONE*, vol. 5, no. 10, p. e13397, 2010.

[134] A. Irrthum, L. Wehenkel, P. Geurts *et al.*, "Inferring regulatory networks from expression data using tree-based methods," *PloS ONE*, vol. 5, no. 9, p. e12776, 2010.

[135] R. Küffner, T. Petri, P. Tavakkolkhah, L. Windhager, and R. Zimmer, "Inferring gene regulatory networks by ANOVA," *Bioinformatics*, vol. 28, no. 10, pp. 1376–1382, 2012.

[136] G. Yang, L. Wang, and X. Wang, "Reconstruction of complex directional networks with group lasso nonlinear conditional Granger causality," *Scientific Reports*, vol. 7, no. 1, p. 2991, 2017.

[137] D. Marbach, T. Schaffter, D. Floreano, R. J. Prill, and G. Stolovitzky, "The DREAM4 in-silico network challenge," *Draft, version 0.3*, 2009.

[138] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, A. Aderhold, R. Bonneau, Y. Chen *et al.*, "Wisdom of crowds for robust gene network inference," *Nature Methods*, vol. 9, no. 8, p. 796, 2012.

[139] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano, "Generating realistic in silico gene networks for performance assessment of reverse engineering methods," *Journal of Computational Biology*, vol. 16, no. 2, pp. 229–239, 2009.

[140] M. Taylor-Teeples, L. Lin, M. De Lucas, G. Turco, T. Toal, A. Gaudinier, N. Young, G. Trabucco, M. Veling, R. Lamothe *et al.*, "An Arabidopsis gene regulatory network for secondary cell wall synthesis," *Nature*, vol. 517, no. 7536, p. 571, 2015.

[141] N. Redekar, G. Pilot, V. Raboy, S. Li, and S. Maroof, "Inference of transcription regulatory network in low phytic acid soybean seeds," *Frontiers in Plant Science*, vol. 8, p. 2029, 2017.

[142] R. M. Shahan, C. Zawora, H. M. Wight, J. Sittmann, W. Wang, S. M. Mount, and Z. Liu, "Consensus co-expression networks provide insights into wild strawberry flower and fruit development," *Plant Physiology*, pp. pp–00 086, 2018.

[143] M. Shibata, C. Breuer, A. Kawamura, N. M. Clark, B. Rymen, L. Braidwood, K. Morohashi, W. Busch, P. N. Benfey, R. Sozzani *et al.*, "GTL1 and DF1 regulate root hair growth through transcriptional repression of ROOT HAIR DEFECTIVE 6-LIKE 4 in Arabidopsis," *Development*, vol. 145, no. 3, p. dev159707, 2018.

[144] T. Hao, D. Wu, L. Zhao, Q. Wang, E. Wang, and J. Sun, "The genome-scale integrated networks in microorganisms," *Frontiers in Microbiology*, vol. 9, p. 296, 2018.

[145] J. M. Lee, E. P. Gianchandani, J. A. Eddy, and J. A. Papin, "Dynamic analysis of integrated signaling, metabolic, and regulatory networks," *PLoS Computational Biology*, vol. 4, no. 5, p. e1000086, 2008.

[146] S. Imam, S. Schäuble, A. N. Brooks, N. S. Baliga, and N. D. Price, "Data-driven integration of genome-scale regulatory and metabolic network models," *Frontiers in Microbiology*, vol. 6, p. 409, 2015.

[147] Z. Wang, S. A. Danziger, B. D. Heavner, S. Ma, J. J. Smith, S. Li, T. Herricks, E. Simeonidis, N. S. Baliga, J. D. Aitchison *et al.*, "Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast," *PLoS Computational Biology*, vol. 13, no. 5, p. e1005489, 2017.

[148] J. A. Suzich, J. F. D. Dean, and K. M. Herrmann, "3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase from carrot root (Daucus carota) is a hysteretic enzyme," *Plant Physiology*, vol. 79, no. 3, pp. 765–770, 1985. [Online]. Available: http://www.plantphysiol.org/content/79/3/765

[149] J. E. B. P. Pinto, J. A. Suzich, and K. M. Herrmann, "3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from potato tuber (Solanum tuberosum L.)," *Plant Physiology*, vol. 82, no. 4, pp. 1040–1044, 1986. [Online]. Available: http://www.plantphysiol.org/content/82/4/1040

[150] W. E. Dyer, J. M. Henstrand, A. K. Handa, and K. M. Herrmann, "Wounding induces the first enzyme of the shikimate pathway in Solanaceae," *Proceedings of the National Academy of Sciences*, vol. 86, no. 19, pp. 7370–7373, 1989. [Online]. Available: https://www.pnas.org/content/86/19/7370

[151] B. Keith, X. N. Dong, F. M. Ausubel, and G. R. Fink, "Differential induction of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase genes in Arabidopsis thaliana by wounding and pathogenic attack," *Proceedings of the National Academy of Sciences*, vol. 88, no. 19, pp. 8821–8825, 1991. [Online]. Available: https://www.pnas.org/content/88/19/8821

[152] A. Carruthers and D. Melchior, "Study of the relationship between bilayer water permeability and bilayer physical state," *Biochemistry*, vol. 22, no. 25, pp. 5797–5807, 1983.

[153] A. C. Chakrabarti and D. W. Deamer, "Permeability of lipid bilayers to amino acids and phosphate," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1111, no. 2, pp. 171 – 177, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005273692903089

[154] R. A. Williams, C. D. Mamotte, and J. R. Burnett, "Phenylketonuria: an inborn error of phenylalanine metabolism," *The Clinical Biochemist Reviews*, vol. 29, no. 1, p. 31, 2008.

[155] G. Fuchs, M. Boll, and J. Heider, "Microbial degradation of aromatic compoundsfrom one strategy to four," *Nature Reviews Microbiology*, vol. 9, no. 11, p. 803, 2011.

[156] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, 2007.

[157] H. Maeda, H. Yoo, and N. Dudareva, "Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate," *Nature Chemical Biology*, vol. 7, no. 1, p. 19, 2011.

[158] J. Pi, P. J. Wookey, and A. J. Pittard, "Cloning and sequencing of the pheP gene, which encodes the phenylalanine-specific transport system of Escherichia coli," *Journal of Bacteriology*, vol. 173, no. 12, pp. 3622–3629, 1991.

[159] J. K. Saini, R. Saini, and L. Tewari, "Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments," *3 Biotech*, vol. 5, no. 4, pp. 337–353, 2015.

[160] S. Kim and B. E. Dale, "Global potential bioethanol production from wasted crops and crop residues," *Biomass-Bioenergy*, vol. 26, no. 4, pp. 361–375, 2004.

[161] H. Zabed, J. Sahu, A. Boyce, and G. Faruq, "Fuel ethanol production from lignocellulosic biomass: an overview on feedstocks and technological approaches," *Renewable and Sustainable Energy Reviews*, vol. 66, pp. 751–774, 2016.

[162] C. Alvarez, F. M. Reyes-Sosa, and B. Díez, "Enzymatic hydrolysis of biomass from wood," *Microbial Biotechnology*, vol. 9, no. 2, pp. 149–156, 2016.

[163] N. Pareek, T. Gillgren, and L. J. Jönsson, "Adsorption of proteins involved in hydrolysis of lignocellulose on lignins and hemicelluloses," *Bioresource Technology*, vol. 148, pp. 70–77, 2013.

[164] T. Huang, T. Tohge, A. Lytovchenko, A. R. Fernie, and G. Jander, "Pleiotropic physiological consequences of feedback-insensitive phenylalanine biosynthesis in Arabidopsis thaliana," *Plant Journal*, vol. 63, no. 5, pp. 823–835, 2010.

[165] M. J. Chen, V. Vijaykumar, B. W. Lu, B. Xia, and N. Li, "Cis- and trans-cinnamic acids have different effects on the catalytic properties of Arabidopsis phenylalanine ammonia lyases PAL1, PAL2, and PAL4," *Journal of Integrative Plant Biology*, vol. 47, no. 1, pp. 67–75, 2005.

[166] G. S. Hu, J. M. Jia, Y. J. Hur, Y. S. Chung, J. H. Lee, D. J. Yun, W. S. Chung, G. H. Yi, T. H. Kim, and D. H. Kim, "Molecular characterization of phenylalanine ammonia lyase gene from Cistanche deserticola," *Molecular Biology Reports*, vol. 38, no. 6, pp. 3741–3750, 2011.

[167] N. Benkeblia, T. Shinano, and M. Osaki, "Metabolite profiling and assessment of metabolome compartmentation of soybean leaves using non-aqueous fractionation and GC-MS analysis," *Metabolomics*, vol. 3, no. 3, pp. 297–305, 2007.

[168] S. Krueger, P. Giavalisco, L. Krall, M. C. Steinhauser, D. Büssis, B. Usadel, U. I. Flügge, A. R. Fernie, L. Willmitzer, and D. Steinhauser, "A topological map of the compartmentalized Arabidopsis thaliana leaf metabolome," *PLoS ONE*, vol. 6, no. 3, 2011.

[169] K. Schallau and B. H. Junker, "Simulating plant metabolic pathways with enzyme-kinetic models," *Plant Physiology*, vol. 152, no. 4, pp. 1763–1771, 2010. [Online]. Available: http://www.plantphysiol.org/content/152/4/1763

[170] M. D. Haunschild, B. Freisleben, R. Takors, and W. Wiechert, "Investigating the dynamic behavior of biochemical networks using model families," *Bioinformatics*, vol. 21, no. 8, pp. 1617–1625, 2005.

[171] Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[172] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference. A Practical Informatio-Theoric Approach.* Springer, 2003.

[173] H. Link, K. Kochanowski, and U. Sauer, "Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo," *Nature Biotechnology*, vol. 31, no. 4, pp. 357–361, 2013. [Online]. Available: http://www.nature.com/doifinder/10.1038/nbt.2489

[174] J. P. Wang, P. P. Naik, H.-C. Chen, R. Shi, C.-Y. Lin, J. Liu, C. M. Shuford, Q. Li, Y.-H. Sun, S. Tunlaya-Anukit, C. M. Williams, D. C. Muddiman, J. J. Ducoste, R. R. Sederoff, and V. L. Chiang, "Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in Populus trichocarpa," *The Plant Cell*, vol. 26, no. 3, pp. 894–914, 2014. [Online]. Available: http://www.plantcell.org/cgi/doi/10.1105/tpc.113.120881

[175] P. Wang, L. Guo, R. Jaini, A. Klempien, R. M. McCoy, J. A. Morgan, N. Dudareva, and C. Chapple, "A ^{13}C isotope labeling method for the measurement of lignin metabolic flux in Arabidopsis stems," *Plant Methods*, vol. 14, no. 1, p. 51, 2018.

[176] F. Lu and J. Ralph, "The DFRC method for lignin analysis. 2. monomers from isolated lignins," *Journal of Agricultural and Food Chemistry*, vol. 46, no. 608, pp. 547–552, 1998.

[177] X. F. Chang, R. Chandra, T. Berleth, and R. P. Beatson, "Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in Arabidopsis thaliana," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 16, pp. 6825–6834, 2008, pMID: 18666780. [Online]. Available: https://doi.org/10.1021/jf800775f

[178] D. Lee, K. Meyer, C. Chapple, and C. J. Douglas, "Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition," *The Plant Cell*, vol. 9, no. 11, pp. 1985–1998, 1997. [Online]. Available: http://www.plantcell.org/cgi/doi/10.1105/tpc.9.11.1985

[179] Y.-H. Su, W. B. Frommer, and U. Ludewig, "Molecular and functional characterization of a family of amino acid transporters from Arabidopsis," *Plant Physiology*, vol. 136, no. 2, pp. 3104–3113, 2004. [Online]. Available: http://www.plantphysiol.org/content/136/2/3104

[180] H. Yang, M. Krebs, Y.-D. Stierhof, and U. Ludewig, "Characterization of the putative amino acid transporter genes AtCAT2, 3 and 4: The tonoplast localized AtCAT2 regulates soluble leaf amino acids," *Journal of Plant Physiology*, vol. 171, no. 8, pp. 594 – 601, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0176161713004616

[181] H. Winter, D. G. Robinson, and H. W. Heldt, "Subcellular volume and metabolic concentrations in barley leaves," *Planta*, vol. 191, pp. 180–190, 1993.

[182] U. Heinig, M. Gutensohn, N. Dudareva, and A. Aharoni, "The challenges of cellular compartmentalization in plant metabolic engineering," *Current Opinion in Biotechnology*, vol. 24, no. 2, pp. 239 – 246, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0958166912002017

[183] D. Winter, B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson, and N. J. Provart, "An electronic fluorescent pictograph browser for exploring and analyzing large-scale biological data sets," *PLOS ONE*, vol. 2, no. 8, pp. 1–12, 08 2007. [Online]. Available: https://doi.org/10.1371/journal.pone.0000718

[184] H. Návarová, F. Bernsdorff, A.-C. Döring, and J. Zeier, "Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity," *The Plant Cell*, vol. 24, no. 12, pp. 5123–5141, 2012. [Online]. Available: http://www.plantcell.org/content/24/12/5123

[185] S. Rasmussen and R. A. Dixon, "Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway," *The Plant Cell*, vol. 11, no. 8, pp. 1537–1551, 1999. [Online]. Available: http://www.plantcell.org/content/11/8/1537

[186] L. Achnine, E. B. Blancaflor, S. Rasmussen, and R. A. Dixon, "Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis," *The Plant Cell*, vol. 16, no. 11, pp. 3098–3109, 2004. [Online]. Available: http://www.plantcell.org/content/16/11/3098

[187] Y. Zeng, M. E. Himmel, and S.-Y. Ding, "Visualizing chemical functionality in plant cell walls," *Biotechnology for Biofuels*, vol. 10, no. 1, p. 263, 2017.

[188] O. M. Terrett and P. Dupree, "Covalent interactions between lignin and hemicelluloses in plant secondary cell walls," *Current Opinion in Biotechnology*, vol. 56, pp. 97–104, 2019.

[189] Q. Liu, L. Luo, and L. Zheng, "Lignins: Biosynthesis and biological functions in plants," *International Journal of Molecular Sciences*, vol. 19, no. 2, p. 335, 2018.

[190] R. Laschimke, "Investigation of the wetting behaviour of natural lignin-a contribution to the cohesion theory of water transport in plants," *Thermochimica Acta*, vol. 151, pp. 35–56, 1989.

[191] B. Prats Mateu, M. T. Hauser, A. Heredia, and N. Gierlinger, "Waterproofing in Arabidopsis: following phenolics and lipids in situ by confocal Raman microscopy," *Frontiers in Chemistry*, vol. 4, p. 10, 2016.

[192] S. Sattler and D. Funnell-Harris, "Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens?" *Frontiers in Plant Science*, vol. 4, p. 70, 2013.

[193] Q. Liu, L. Zheng, F. He, F.-J. Zhao, Z. Shen, and L. Zheng, "Transcriptional and physiological analyses identify a regulatory role for hydrogen peroxide in the lignin biosynthesis of copper-stressed rice roots," *Plant and Soil*, vol. 387, no. 1-2, pp. 323–336, 2015.

[194] D. Van de Wouwer, W. Boerjan, and B. Vanholme, "Plant cell wall sugars: sweeteners for a bio-based economy," *Physiologia Plantarum*, vol. 164, no. 1, pp. 27–44, 2018.

[195] R. Vanholme, J. Ralph, T. Akiyama, F. Lu, J. R. Pazo, H. Kim, J. H. Christensen, B. Van Reusel, V. Storme, R. De Rycke, A. Rohde, K. Morreel, and W. Boerjan, "Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in Arabidopsis," *The Plant Journal*, vol. 64, no. 6, pp. 885–897, 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2010.04353.x

[196] F. Chen, Y. Tobimatsu, D. Havkin-Frenkel, R. A. Dixon, and J. Ralph, "A polymer of caffeyl alcohol in plant seeds," *Proceedings of the National Academy of Sciences*, vol. 109, no. 5, pp. 1772–1777, 2012. [Online]. Available: https://www.pnas.org/content/109/5/1772

[197] P. A. Saa and L. K. Nielsen, "Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks," *Biotechnology Advances*, vol. 35, no. 8, pp. 981 – 1003, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0734975017301167

[198] X.-G. Zhu, Y. Wang, A. Brutigam, and A. P. M. Weber, "Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis," *Journal of Experimental Botany*, vol. 65, no. 13, pp. 3567–3578, 03 2014. [Online]. Available: https://doi.org/10.1093/jxb/eru058

[199] A. Khodayari and C. D. Maranas, "A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains," *Nature Communications*, vol. 7, p. 13806, 2016.

[200] M. Cloutier, J. Chen, C. De Dobbeleer, M. Perrier, and M. Jolicoeur, "A systems approach to plant bioprocess optimization," *Plant Biotechnology Journal*, vol. 7, no. 9, pp. 939–951, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7652.2009.00455.x

[201] A. Valancin, B. Srinivasan, J. Rivoal, and M. Jolicoeur, "Analyzing the effect of decreasing cytosolic triosephosphate isomerase on Solanum tuberosum hairy root cells using a kineticmetabolic model," *Biotechnology and Bioengineering*, vol. 110, no. 3, pp. 924–935, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.24747

[202] X. Chen, A. P. Alonso, and Y. Shachar-Hill, "Dynamic metabolic flux analysis of plant cell wall synthesis," *Metabolic Engineering*, vol. 18, pp. 78 – 85, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1096717613000487

[203] R. Zhong, C. Lee, and Z.-H. Ye, "Global analysis of direct targets of secondary wall NAC master switches in Arabidopsis," *Molecular Plant*, vol. 3, no. 6, pp. 1087 – 1103, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1674205214605566

[204] R. L. McCarthy, R. Zhong, and Z.-H. Ye, "MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis," *Plant and Cell Physiology*, vol. 50, no. 11, pp. 1950–1964, 2009.

[205] J. I. Kim, W. L. Dolan, N. A. Anderson, and C. Chapple, "Indole glucosinolate biosynthesis limits phenylpropanoid accumulation in Arabidopsis thaliana," *The Plant Cell*, vol. 27, no. 5, pp. 1529–1546, 2015. [Online]. Available: http://www.plantcell.org/content/27/5/1529

[206] W. Deng, C. Gunasekara, H. Wei, K. Zhang, and L. Brown, "TGMI: an efficient algorithm for identifying pathway regulators through evaluation of triple-gene mutual interaction," *Nucleic Acids Research*, vol. 46, no. 11, pp. e67–e67, 03 2018. [Online]. Available: https://doi.org/10.1093/nar/gky210

[207] C. C. Chapple, T. Vogt, B. E. Ellis, and C. R. Somerville, "An Arabidopsis mutant defective in the general phenylpropanoid pathway." *The Plant Cell*, vol. 4, no. 11, pp. 1413–1424, 1992. [Online]. Available: http://www.plantcell.org/content/4/11/1413

[208] M. Ruegger and C. Chapple, "Mutations that reduce sinapoylmalate accumulation in Arabidopsis thaliana define loci with diverse roles in phenylpropanoid metabolism," *Genetics*, vol. 159, no. 4, pp. 1741–1749, 2001. [Online]. Available: http://www.genetics.org/content/159/4/1741

[209] A. L. Schilmiller, J. Stout, J.-K. Weng, J. Humphreys, M. O. Ruegger, and C. Chapple, "Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis," *The Plant Journal*, vol. 60, no. 5, pp. 771–782, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2009.03996.x

[210] J.-K. Weng, T. Akiyama, N. D. Bonawitz, X. Li, J. Ralph, and C. Chapple, "Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants," *The Plant Cell*, vol. 22, no. 4, pp. 1033–1045, 2010. [Online]. Available: http://www.plantcell.org/content/22/4/1033

[211] P. Wang, *Measuring and modeling of phenylpropanoid metabolic flux in Arabidopsis.* Purdue Univerisity Dissertation, 2019.

[212] D. Kim, B. Langmead, and S. L. Salzberg, "HISAT: a fast spliced aligner with low memory requirements," *Nature Methods*, vol. 12, no. 4, p. 357, 2015.

[213] S. Anders, P. T. Pyl, and W. Huber, "HTSeqa Python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, pp. 166–169, 2015.

[214] A. Klempien, Y. Kaminaga, A. Qualley, D. A. Nagegowda, J. R. Widhalm, I. Orlova, A. K. Shasany, G. Taguchi, C. M. Kish, B. R. Cooper *et al.*, "Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers," *The Plant Cell*, vol. 24, no. 5, pp. 2015–2030, 2012.

[215] R. S. Sorenson, M. J. Deshotel, K. Johnson, F. R. Adler, and L. E. Sieburth, "Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and redundancy," *Proceedings of the National Academy of Sciences*, vol. 115, no. 7, pp. E1485–E1494, 2018.

[216] A. Ebrahim, E. Brunk, J. Tan, E. J. O'brien, D. Kim, R. Szubin, J. A. Lerman, A. Lechner, A. Sastry, A. Bordbar *et al.*, "Multi-omic data integration enables discovery of hidden biological regularities," *Nature Communications*, vol. 7, p. 13091, 2016.

[217] A. Cornish-Bowden, *Fundamentals of enzyme kinetics*. Wiley-Blackwell Weinheim, Germany, 2012, vol. 510.

[218] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo, "eQuilibratorthe biochemical thermodynamics calculator," *Nucleic Acids Research*, vol. 40, no. D1, pp. D770–D775, 2011.

[219] E. Noor, A. Bar-Even, A. Flamholz, Y. Lubling, D. Davidi, and R. Milo, "An integrated open framework for thermodynamics of reactions that combines accuracy and coverage," *Bioinformatics*, vol. 28, no. 15, pp. 2037–2044, 2012.

[220] J. Zhou, C. Lee, R. Zhong, and Z.-H. Ye, "MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis," *The Plant Cell*, vol. 21, no. 1, pp. 248–266, 2009. [Online]. Available: http://www.plantcell.org/content/21/1/248

[221] H. Jin, E. Cominelli, P. Bailey, A. Parr, F. Mehrtens, J. Jones, C. Tonelli, B. Weisshaar, and C. Martin, "Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis," *The EMBO Journal*, vol. 19, no. 22, pp. 6150–6161, 2000. [Online]. Available: http://emboj.embopress.org/content/19/22/6150

[222] T. Soyano, S. Thitamadee, Y. Machida, and N.-H. Chua, "ASYMMETRIC LEAVES2-LIKE19/LATERAL ORGAN BOUNDARIES DOMAIN30 and ASL20/LBD18 regulate tracheary element differentiation in Arabidopsis," *The Plant Cell*, vol. 20, no. 12, pp. 3359–3373, 2008. [Online]. Available: http://www.plantcell.org/content/20/12/3359

[223] R. Vanholme, V. Storme, B. Vanholme, L. Sundin, J. H. Christensen, G. Goeminne, C. Halpin, A. Rohde, K. Morreel, and W. Boerjan, "A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis," *The Plant Cell*, vol. 24, no. 9, pp. 3506–3529, 2012.

[224] L. G. Landry, C. C. Chapple, and R. L. Last, "Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage," *Plant Physiology*, vol. 109, no. 4, pp. 1159–1166, 1995.

[225] X. Li, N. D. Bonawitz, J.-K. Weng, and C. Chapple, "The growth reduction associated with repressed lignin biosynthesis in Arabidopsis thaliana is independent of flavonoids," *The Plant Cell*, vol. 22, no. 5, pp. 1620–1632, 2010.

[226] D. Öhman, B. Demedts, M. Kumar, L. Gerber, A. Gorzsás, G. Goeminne, M. Hedenström, B. Ellis, W. Boerjan, and B. Sundberg, "MYB103 is required for ferulate-5-hydroxylase expression and syringyl lignin biosynthesis in Arabidopsis stems," *The Plant Journal*, vol. 73, no. 1, pp. 63–76, 2013.

[227] S. G. Hussey, E. Mizrachi, A. V. Spokevicius, G. Bossinger, D. K. Berger, and A. A. Myburg, "SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in Arabidopsis fibres and increases fibre cell area in Eucalyptus," *BMC Plant Biology*, vol. 11, no. 1, p. 173, Dec 2011. [Online]. Available: https://doi.org/10.1186/1471-2229-11-173

[228] H. Shen, M. Mazarei, H. Hisano, L. Escamilla-Trevino, C. Fu, Y. Pu, M. R. Rudis, Y. Tang, X. Xiao, L. Jackson *et al.*, "A genomics approach to deciphering lignin biosynthesis in switchgrass," *The Plant Cell*, vol. 25, no. 11, pp. 4342–4361, 2013.

[229] R. B. Nair, K. L. Bastress, M. O. Ruegger, J. W. Denault, and C. Chapple, "The Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis," *The Plant Cell*, vol. 16, no. 2, pp. 544–554, 2004.

[230] J. M. Rohwer and F. C. Botha, "Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data," *Biochemical Journal*, vol. 358, no. 2, pp. 437–445, 2001. [Online]. Available: http://www.biochemj.org/content/358/2/437

[231] G. Curien, S. Ravanel, and R. Dumas, "A kinetic model of the branch-point between the methionine and threonine biosynthesis pathways in Arabidopsis thaliana," *European Journal of Biochemistry*, vol. 270, no. 23, pp. 4615–4627, 2003. [Online]. Available: https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1432-1033.2003.03851.x

[232] A. Laisk, H. Eichelmann, and V. Oja, "C3 photosynthesis in silico," *Photosynthesis Research*, vol. 90, no. 1, pp. 45–66, Oct 2006. [Online]. Available: https://doi.org/10.1007/s11120-006-9109-1

[233] R. Rios-Estepa, G. W. Turner, J. M. Lee, R. B. Croteau, and B. M. Lange, "A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint," *Proceedings of the National Academy of Sciences*, vol. 105, no. 8, pp. 2818–2823, 2008. [Online]. Available: https://www.pnas.org/content/105/8/2818

[234] M. Cloutier, J. Chen, F. Tatge, V. McMurray-Beaulieu, M. Perrier, and M. Jolicoeur, "Kinetic metabolic modelling for the control of plant cells cytoplasmic phosphate," *Journal of Theoretical Biology*, vol. 259, no. 1, pp. 118 – 131, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022519309000976

[235] H. Chen, H. Jiang, and J. A. Morgan, "Non-natural cinnamic acid derivatives as substrates of cinnamate 4-hydroxylase," *Phytochemistry*, vol. 68, no. 3, pp. 306–311, 2007.

[236] M. Pietrowska-Borek, H.-P. Stuible, E. Kombrink, and A. Guranowski, "4-coumarate: coenzyme a ligase has the catalytic capacity to synthesize and reuse various (di) adenosine polyphosphates," *Plant Physiology*, vol. 131, no. 3, pp. 1401–1410, 2003.

[237] G. Schoch, S. Goepfert, M. Morant, A. Hehn, D. Meyer, P. Ullmann, and D. Werck-Reichhart, "CYP98A3 from Arabidopsis thaliana is a 3-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway," *Journal of Biological Chemistry*, vol. 276, no. 39, pp. 36 566–36 574, 2001.

APPENDICES

A. SUPPLEMENTARY TABLE FOR INTRODUCTION

Table A.1.: List of kinetic models in plants. A complete list of plant kinetic models is shown. Models presented in [82] are not included here.

Organism	Tissue/Cell Type	Pathway	Compartments	Rate Equations	Parameter Assignments	References
Saccharum officinarum	Culm Tissue from Internode 3-10	Sucrose Metabolism	Cytoplasm	Michaelis-Menten; Hill Equation	Measured	[96, 230]
Arabidopsis thaliana	Leaf Mesophyll Cells	Aspartate-Derived Amino-Acid Pathway	Chloroplast	Michaelis-Menten; Mass Action; Hill Equation	Measured	[91, 231]
Nicotiana tabacum L.	Leaves	Photosynthesis	Cytoplasm; Chloroplast; Thylakoid Membranes and Lumen	Michaelis-Menten	Measured; Fitted	[232]
Catharanthus roseus	Hairy Roots	Central Carbon Metabolism	Not Specified	Michaelis-Menten; Hill Equation	Measured; Fitted	[104]
Solanum tuberosum	Tubers	Phenylpropanoid Metabolism	Cytoplasm	Mass Action; Generalized Mass Action	Fitted	[100]
"Typical" C3 Plant	C3 Leaves	Photosynthetic C3 Carbon Metabolism	Cytoplasm; Chloroplast	Michaelis-Menten	Measured	[97]
Arabidopsis thaliana	Leaves	Aliphatic Glucosinolate Biosynthesis	Not Specified	Michaelis-Menten; Mass Action	Measured; Fitted	[116]
Mentha piperita	Secretory Cells in Glandular Trichomes	*p*-Menthane Monoterpene Biosynthesis	Cytoplasm	Michaelis-Menten	Measured; Fitted	[114, 233]
Eschscholtiza californica	Cell Cultures	Central Carbon Metabolism	Cytoplasm; Vacuole	Michaelis-Menten; Hill Equation	Fitted	[200, 234]
Arabidopsis thaliana	Cell Cultures	Fenclorim Metabolism	Cytoplasm; Cell Wall	Michaelis-Menten; Mass Action	Measured; Fitted	[101]
Arabidopsis thaliana	Leaves	Flavonoid Pathway	Not Specified	Mass Action	Fitted	[102]
Petunia hybrida	Flowers	Benzenoid Network	Cytoplasm	Michaelis-Menten; Mass Action	Fitted	[103]
Populus trichocarpa	Xylem	Lignin Biosynthesis	Cytoplasm	Generalized Mass Action	Fitted	[92]
Arabidopsis thaliana	Leaves	Central Carbohydrate Metabolism	Not Specified	Michaelis-Menten	Fitted	[105]
Medicago Sativa L.	Sixth Internode	Lignin Biosynthesis	Cytoplasm	Generalized Mass Action	Fitted	[93]
Arabidopsis thaliana	Cell Cultures	Cell Wall Biosynthesis	Not Specified	Mass Action	Fitted	[202]
Solanum tuberosum	Hairy Roots	Central Carbon Metabolism	Not Specified	Michaelis-Menten	Measured; Fitted	[201]

continued on next page

Table A.1.: *continued*

Organism	Tissue/Cell Type	Pathway	Compartments	Rate Equations	Parameter Assignments	References
Solanum lycopersicum	Pericarp	Sugar Metabolism	Cytoplasm; Plastid; Vacuole	Michaelis-Menten; Mass Action; Hill Equation	Measured; Fitted	[115]
Populus trichocarpa	Secondary Differentiating Xylem	Lignin Biosynthesis	Cytoplasm	Michaelis-Menten	Measured	[174]
C4 Plants	Bundle Sheath and Mesophyll Cells	Photosynthesis	Cytoplasm; Chloroplast; Peroxisome	Michaelis-Menten	Measured	[98, 198]
Panicum virgatum	Stem and Tiller Tissue	Lignin Biosynthesis	Cytoplasm	Generalized Mass Action	Fitted	[94]
Brachypodium distachyon	Not Specified	Lignin Biosynthesis	Cytoplasm; ; Outer Endoplasmic Reticulum Surface	Generalized Mass Action	Fitted	[95]
Arabidopsis thaliana	Basal Stem	Phenylpropanoid Metabolism	Cytoplasm; Plastid; Vacuole	Michaelis-Menten	Fitted	[109]

B. SUPPLEMENTARY FIGURES AND TABLE FOR DYNAMIC MODELING OF SUBCELLULAR PHENYLPROPANOID METABOLISM IN ARABIDOPSIS LIGNIFYING CELLS

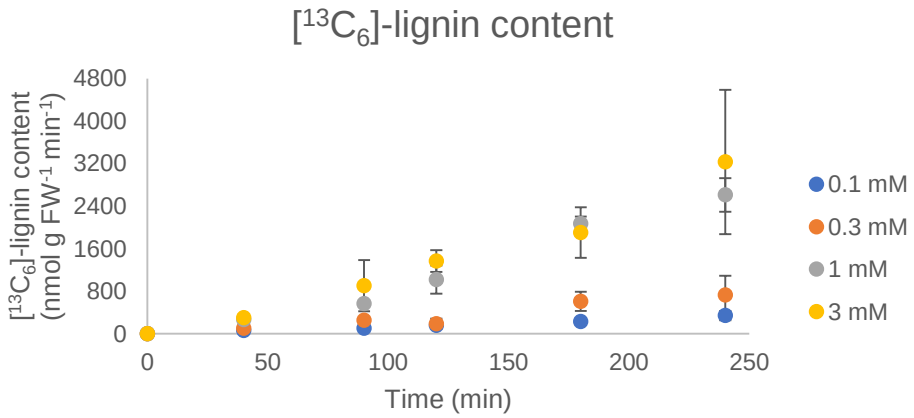

Fig. B.1. Total [$^{13}C_6$]-lignin in stems fed with [$^{13}C_6$]-Phe. Measurements are represented as points by average ± standard deviation (n=3).

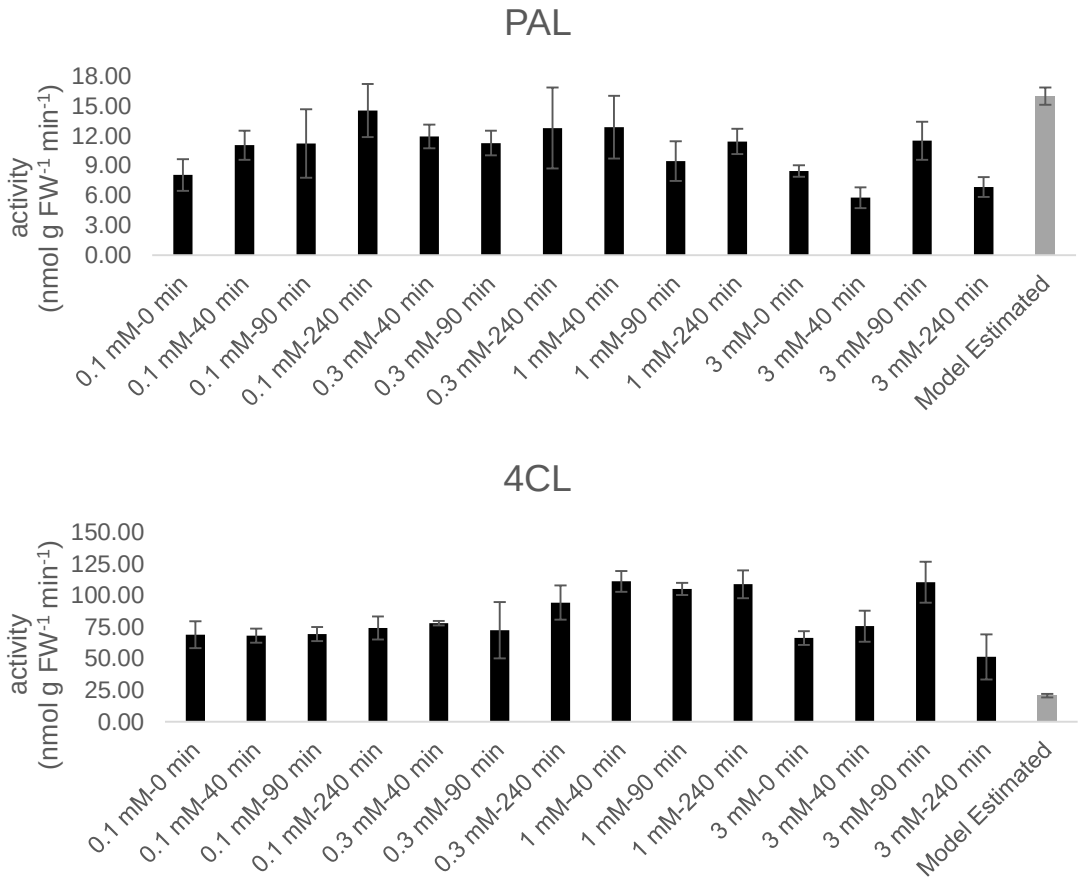

Fig. B.2. Measured and estimated PAL and 4CL activities in wild type stems over the feeding period. Shown are the maximal capacities of PAL and 4CL, which were obtained from measurements within the feeding period (black bars), or refined model prediction (grey bars). All bars are presented as average ± standard deviation (n=3 for measurements).

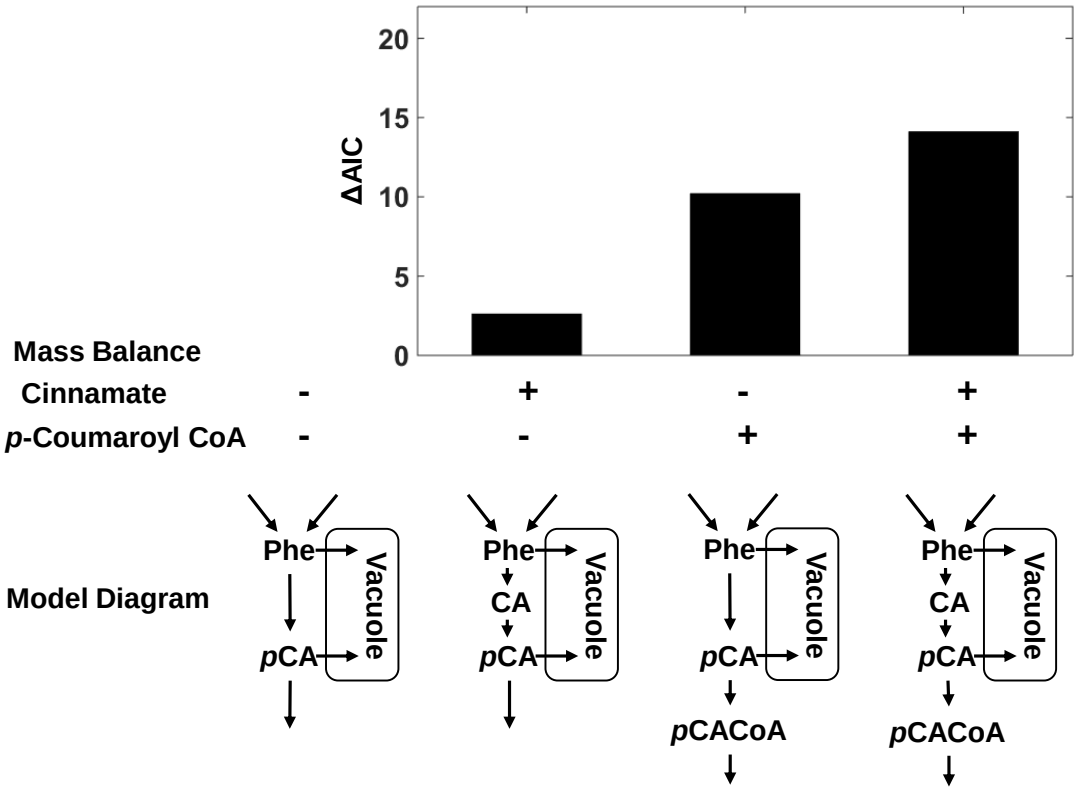

Fig. B.3. AIC scores of the models relative to one without cinnamate and *p*-coumaroyl CoA mass balances. Three additional models were generated by considering all possible combinations of mass balances of cinnamate and *p*-coumaroyl CoA. Corrected AIC was applied to evaluate the performance of these models. Abbreviations: CA, cinnamate; *p*CACoA, *p*-coumaroyl CoA.

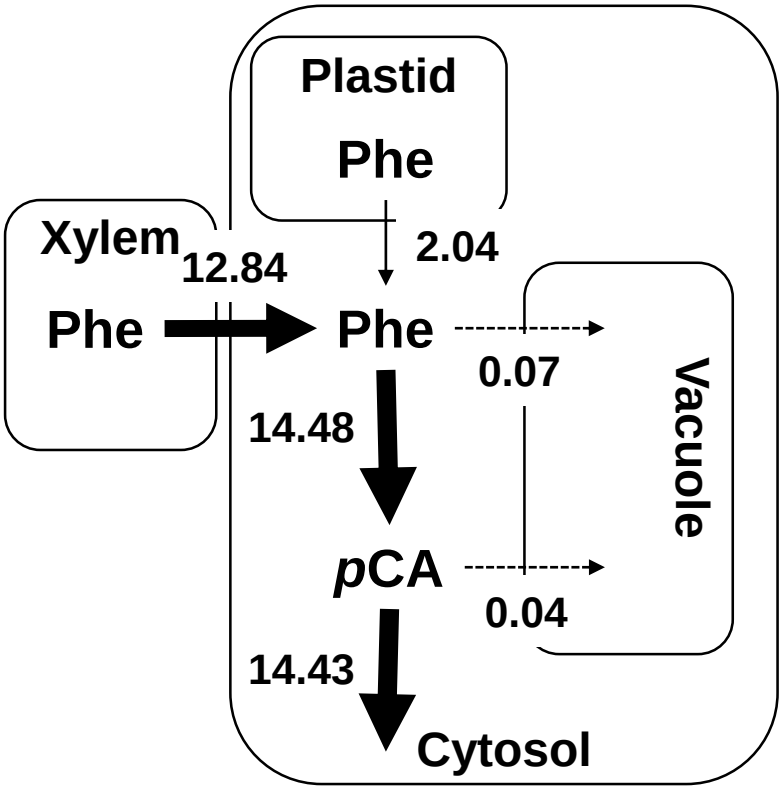

Fig. B.4. Flux distribution of the phenylpropanoid pathway during 3 mM [$^{13}C_6$]-Phe feeding condition. Shown were the fluxes at 240 min after feeding, predicted by the refined model. The width of the arrows are proportional to the flux through the reactions (Units: nmol g FW^{-1} min^{-1}).

Table B.1.
AIC scores of the models relative to the one without any potential metabolite-enzyme interactions. Shown are the relative change of AIC scores when each interaction was introduced into the refined model. No interaction significantly improves the models performance since all AIC score changes are positive.

Introduced interaction	ΔAIC
Phe_v competitively inhibits VCAT	0.32
Phe_c competitively inhibits PXT	0.33
Phe_v uncompetitively inhibits VCAT	0.46
$p\text{CA}_c$ uncompetitively inhibits PXT	1.57
Phe_c uncompetitively inhibits PCAT	2.25
$p\text{CA}_c$ uncompetitively inhibits PAL	2.25
$p\text{CA}_v$ uncompetitively inhibits PVT	2.59
$p\text{CA}_c$ competitively inhibits PAL	2.68
$p\text{CA}_c$ uncompetitively inhibits PCAT	2.69
$p\text{CA}_v$ competitively inhibits PVT	2.74
$p\text{CA}_c$ competitively inhibits PXT	2.74
$p\text{CA}_c$ competitively inhibits PCAT	2.76
Phe_c activates 4CL	2.79
Phe_c uncompetitively inhibits PXT	18.04

C. SUPPLEMENTARY FIGURES AND TABLES FOR COMBINING GENE REGULATORY NETWORK AND KINETIC MODELING OF LIGNIN BIOSYNTHESIS IN ARABIDOPSIS

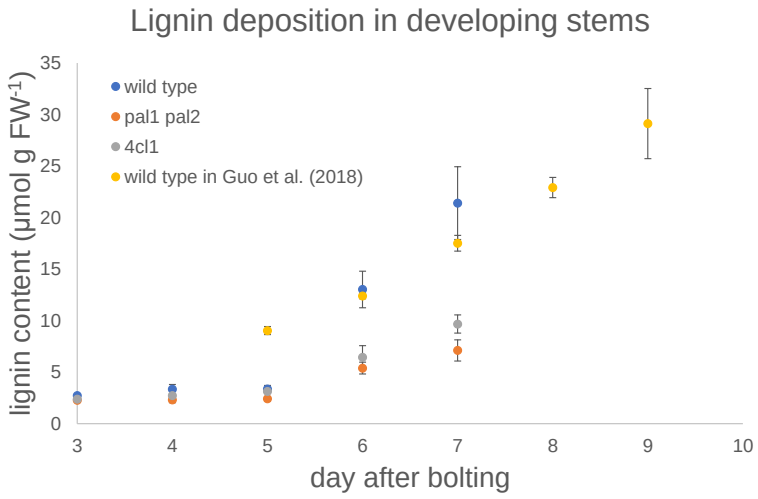

Fig. C.1. Lignin deposition in wild type, *pal1 pal2* and *4cl1* Arabidopsis. Molecular weight for lignin monomer was assumed to be 202.4 g/mol, which was calculated based on wild type ratios of H, G and S lignin and their molecular weights respectively.

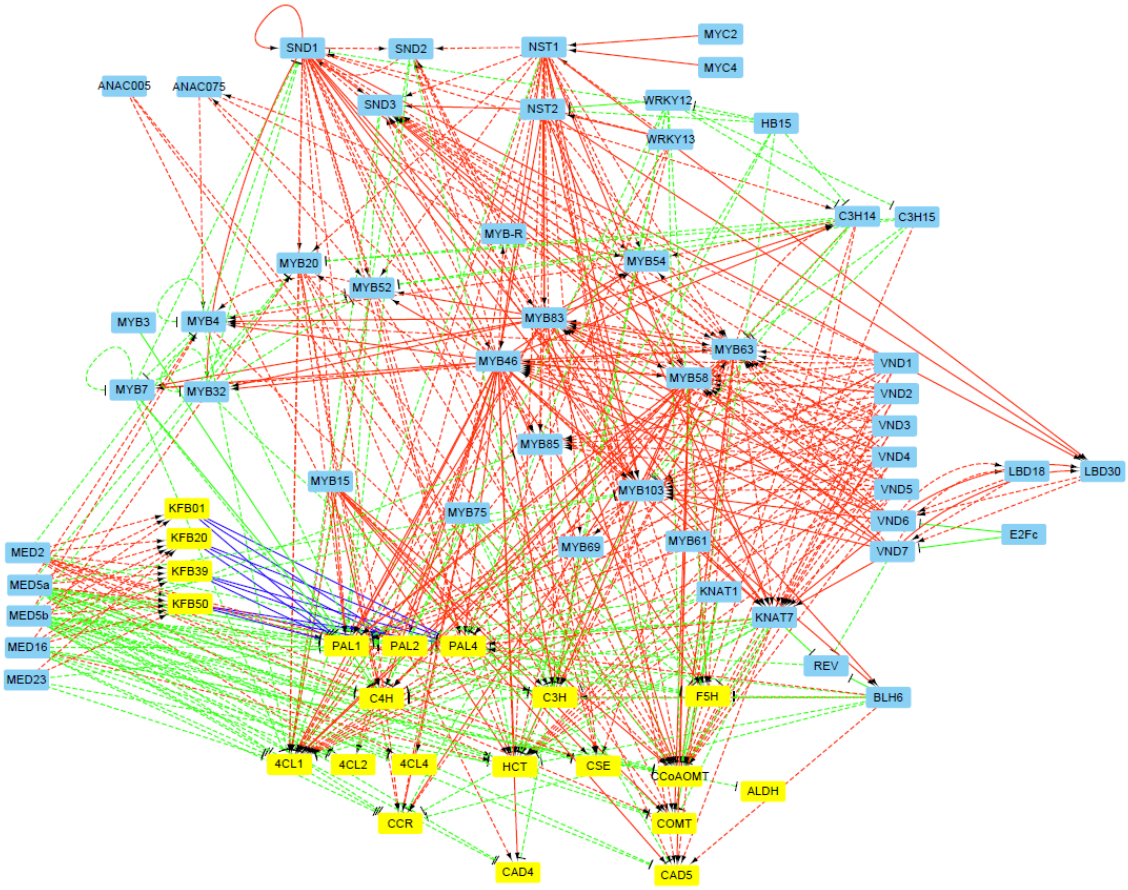

Fig. C.2. Reported regulatory interactions of lignin biosynthesis in Arabidopsis. Each node in the network represents one gene involved in lignin biosynthesis, with yellow color for structural genes and blue for TFs. Interactions are represented as arrows between the TFs and the targets. Activations are shown in red and inhibitions are shown in green. Arrows in solid lines represents the corresponding interactions were supported with evidence of direct binding, while arrows in dashed lines represents the interactions without direct binding evidence. The plot was generated with Cytoscape 3.7.0.

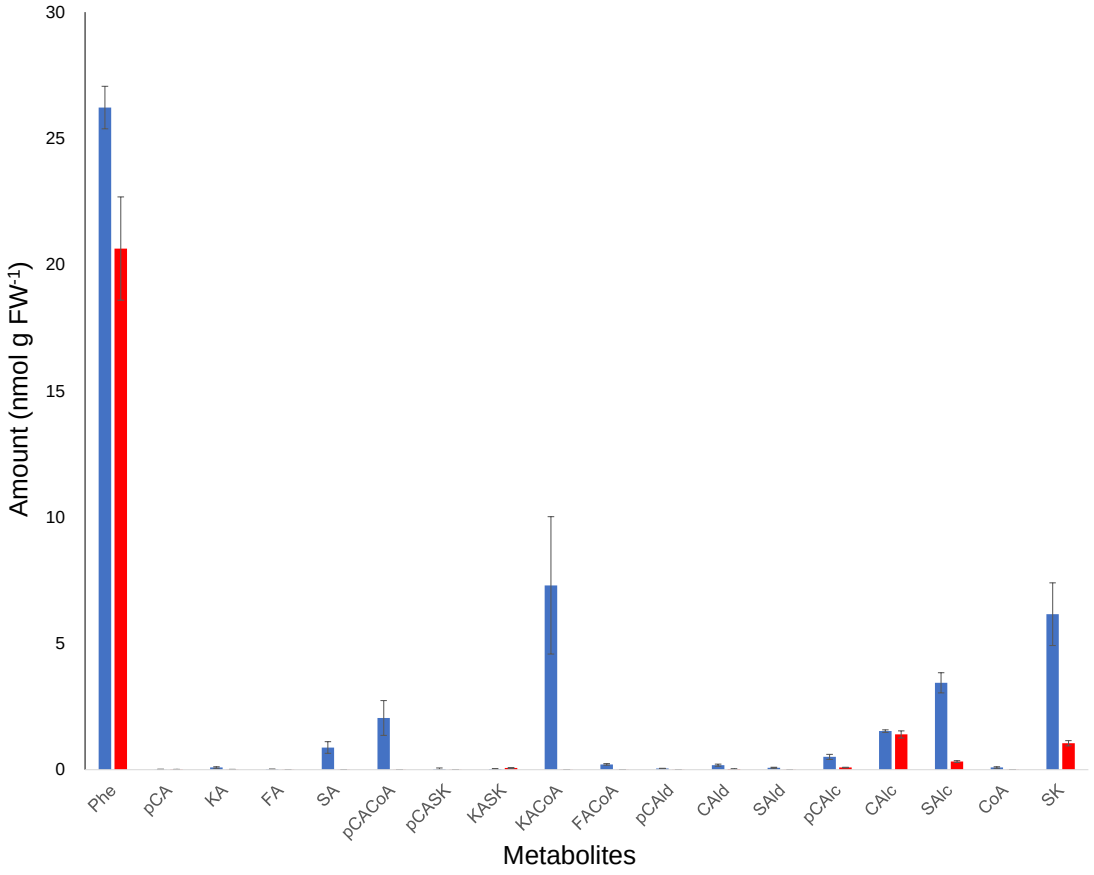

Fig. C.3. Metabolite concentrations of lignin biosynthesis in wild type Arabidopsis. Blue bars represent LC-MS/MS measurements for wild type Arabidopsis, and red bars represent predicted metabolite concentrations from the hierarchical model. The predicted values were obtained once the model reached a steady state. Bars for measurements are represented as average ± standard deviation. Bars for model predictions are shown as point estimation ± standard deviation, with the latter calculated from 1,000 MCMC samples.

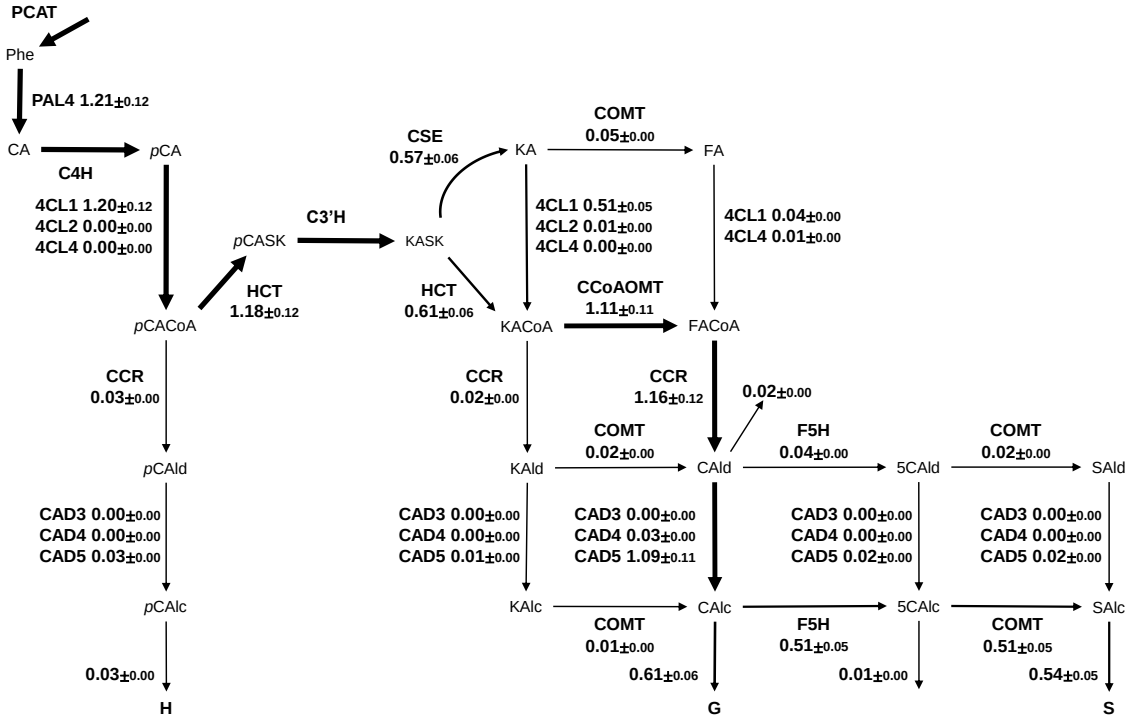

Fig. C.4. Flux map of lignin biosynthesis in *pal1 pal2* Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

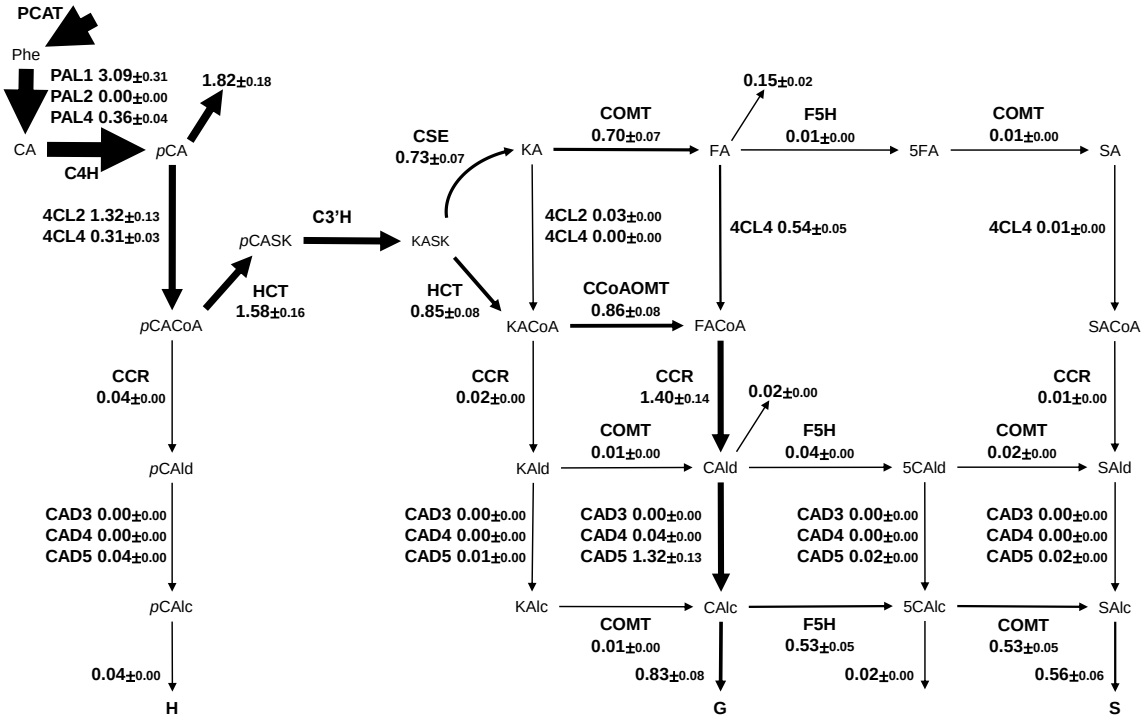

Fig. C.5. Flux map of lignin biosynthesis in *4cl1* Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

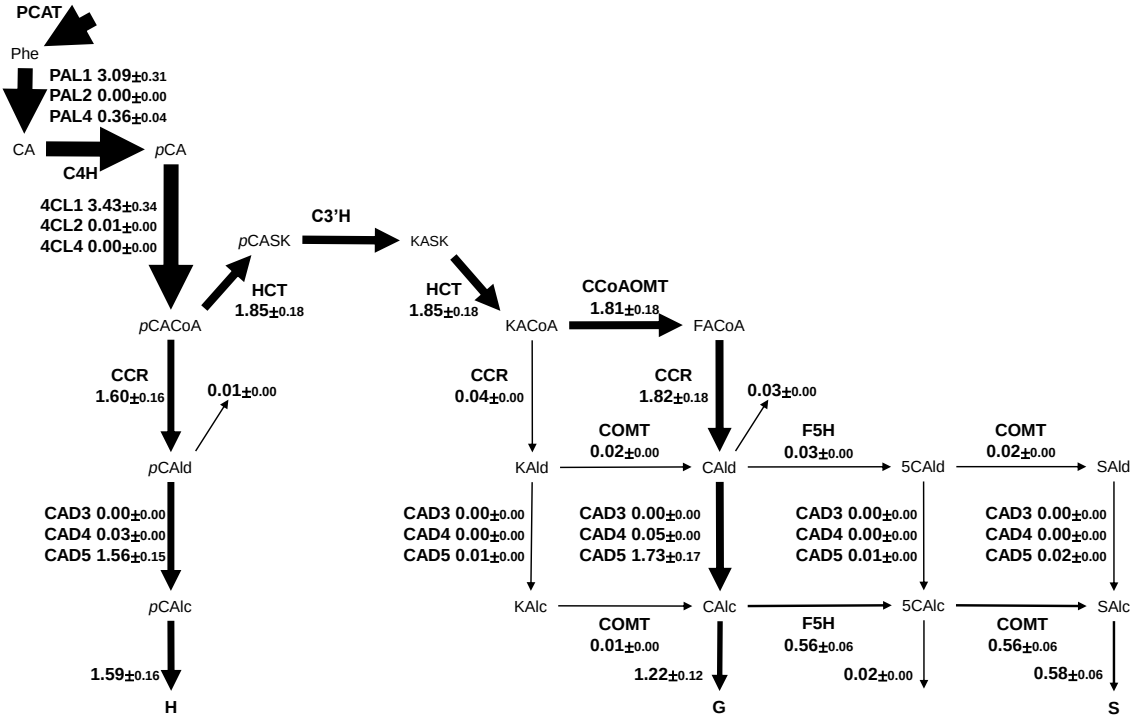

Fig. C.6. Flux map of lignin biosynthesis in *cse-2* Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

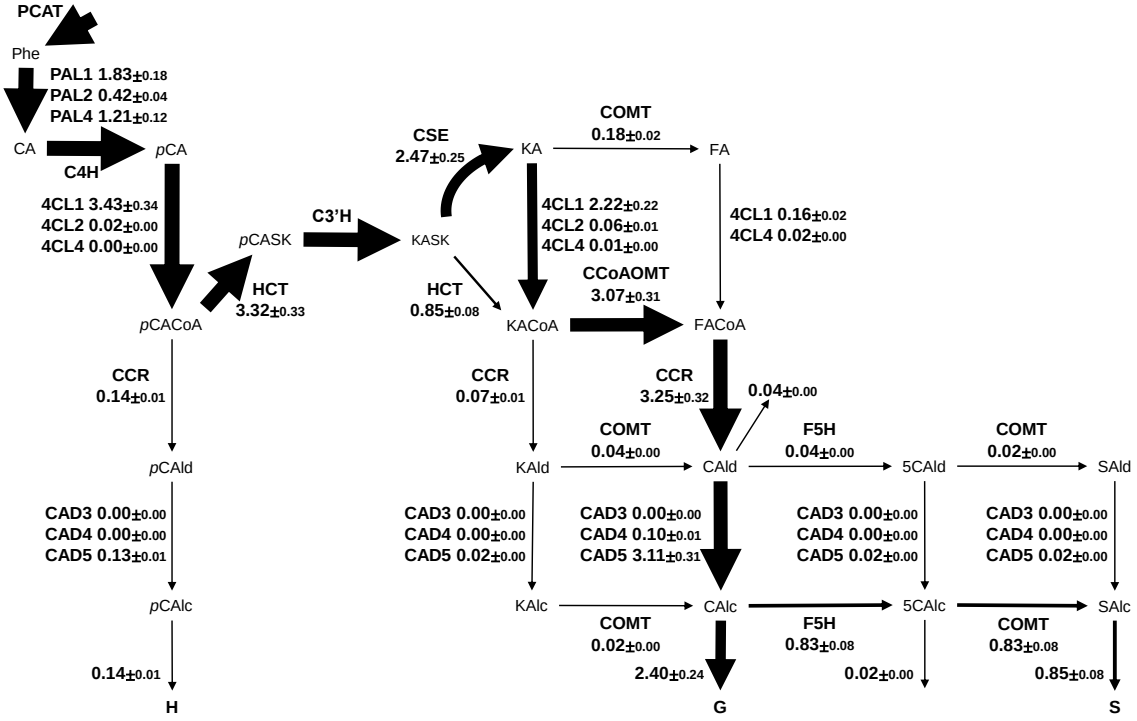

Fig. C.7. Flux map of lignin biosynthesis in *med5a/b* Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

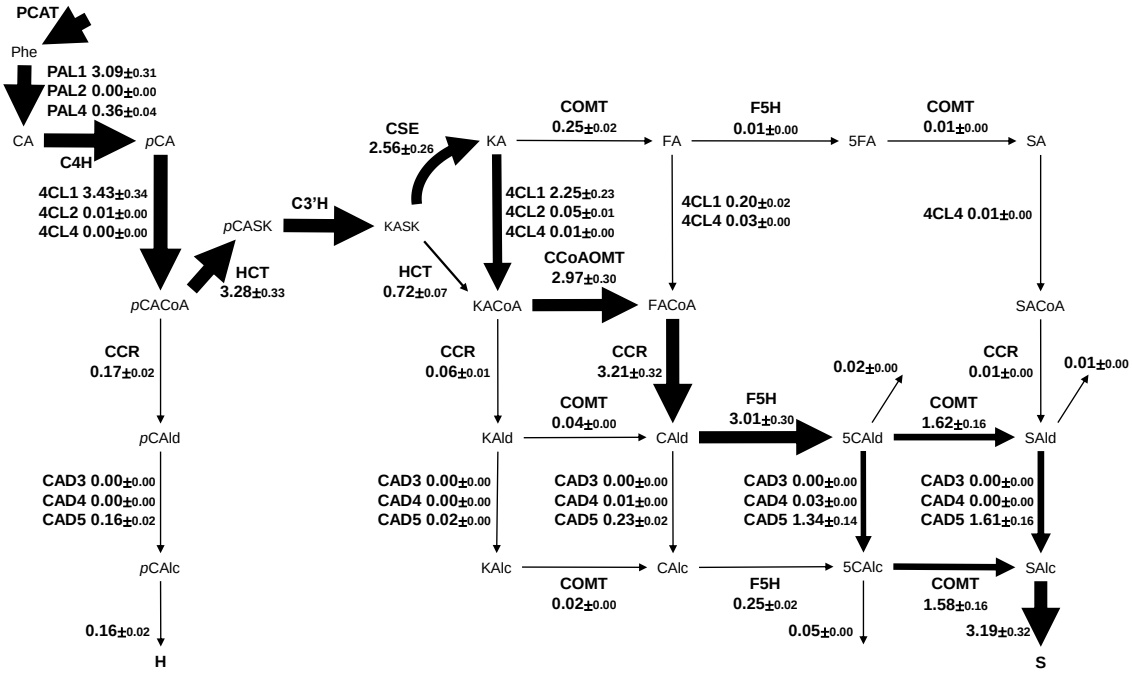

Fig. C.8. Flux map of lignin biosynthesis in *F5H*-overexpressing Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

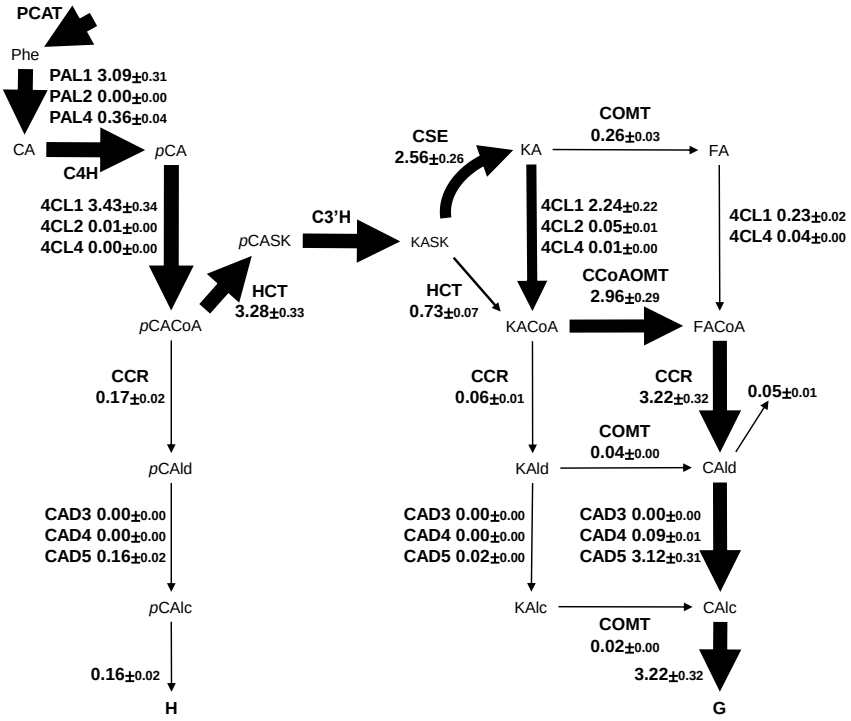

Fig. C.9. Flux map of lignin biosynthesis in *fah1* Arabidopsis from the hierarchical model (Unit: nmol g FW^{-1} min^{-1}).

Table C.1.: List of parameters collected from the literature for kinetic modeling.

Parameter Type	Enzyme	Metabolite	Value	Reference
V_{max}	PAL1	Phe	5.5 pkat/μg protein	[34]
V_{max}	PAL2	Phe	10.5 pkat/μg protein	[34]
V_{max}	PAL4	Phe	9.9 pkat/μg protein	[34]
K_m	PAL1	Phe	68 μM	[34]
K_m	PAL2	Phe	64 μM	[34]
K_m	PAL4	Phe	71 μM	[34]
k_{cat}	C4H	CA	102.9 min^{-1}	[235]
K_m	C4H	CA	0.5 μM	[235]
V_{max}	4CL1	*p*CA	1251.1 nkat/mg protein[a]	[29]
V_{max}	4CL1	KA	337.8 nkat/mg protein[a]	[29]
V_{max}	4CL1	FA	663.1 nkat/mg protein[a]	[29]
K_m	4CL1	*p*CA	38 μM	[29]
K_m	4CL1	KA	11 μM	[29]
K_m	4CL1	FA	199 μM	[29]
V_{max}	4CL2	*p*CA	310 nkat/mg protein	[236]
V_{max}	4CL2	KA	158 nkat/mg protein	[236]
K_m	4CL2	*p*CA	233 μM	[236]
K_m	4CL2	KA	24 μM	[236]
V_{max}	4CL4	*p*CA	100 nkat/mg protein	[30]
V_{max}	4CL4	KA	187 nkat/mg protein	[30]
V_{max}	4CL4	FA	153 nkat/mg protein	[30]
V_{max}	4CL4	SA	105 nkat/mg protein	[30]
K_m	4CL4	*p*CA	432 μM	[30]
K_m	4CL4	KA	186 μM	[30]
K_m	4CL4	FA	26 μM	[30]
K_m	4CL4	SA	20 μM	[30]

continued on next page

Table C.1.: *continued*

Parameter Type	Enzyme	Metabolite	Value	Reference
V_{max}	CSE	KASK	9.3 pkat/μg protein	[21]
K_m	CSE	KASK	96.5 μM	[21]
k_{cat}	C3'H	*p*CASK	612 min^{-1}	[237]
K_m	C3'H	*p*CASK	7 μM	[237]
k_{cat}	CCR	*p*CACoA	1.63 min^{-1}	[37]
k_{cat}	CCR	KACoA	2.18 min^{-1}	[36]
k_{cat}	CCR	FACoA	1.9 min^{-1}	[37]
k_{cat}	CCR	SACoA	1.67 min^{-1}	[37]
K_m	CCR	*p*CACoA	2.27 μM	[37]
K_m	CCR	KACoA	12.5 μM	[36]
K_m	CCR	FACoA	0.42 μM	[37]
K_m	CCR	SACoA	0.55 μM	[37]
V_{max}	COMT	KA	14.6 pkat/μg protein	[41]
V_{max}	COMT	5FA	30.1 pkat/μg protein	[41]
V_{max}	COMT	KAld	35.9 pkat/μg protein	[41]
V_{max}	COMT	5CAld	66.2 pkat/μg protein	[41]
V_{max}	COMT	KAlc	35.3 pkat/μg protein	[41]
V_{max}	COMT	5CAlc	51.9 pkat/μg protein	[41]
K_m	COMT	KA	24.2 μM	[41]
K_m	COMT	5FA	32.0 μM	[41]
K_m	COMT	KAld	19.7 μM	[41]
K_m	COMT	5CAld	17.9 μM	[41]
K_m	COMT	KAlc	51.5 μM	[41]
K_m	COMT	5CAlc	31.6 μM	[41]
V_{max}	F5H	FA	4 pkat/mg protein	[40]
V_{max}	F5H	CAld	5 pkat/mg protein	[40]
V_{max}	F5H	CAlc	6 pkat/mg protein	[40]

continued on next page

Table C.1.: *continued*

Parameter Type	Enzyme	Metabolite	Value	Reference
K_m	F5H	FA	1000 μM	[40]
K_m	F5H	CAld	1 μM	[40]
K_m	F5H	CAlc	3 μM	[40]
k_{cat}	CAD5	*p*CAld	14.52 s^{-1}	[33]
k_{cat}	CAD5	KAld	7.3 s^{-1}	[33]
k_{cat}	CAD5	CAld	12.19 s^{-1}	[33]
k_{cat}	CAD5	5CAld	8.28 s^{-1}	[33]
k_{cat}	CAD5	SAld	13.72 s^{-1}	[33]
K_m	CAD5	*p*CAld	13 μM	[33]
K_m	CAD5	KAld	68 μM	[33]
K_m	CAD5	CAld	35 μM	[33]
K_m	CAD5	5CAld	22 μM	[33]
K_m	CAD5	SAld	20 μM	[33]
k_{cat}	CAD4	*p*CAld	3.44 s^{-1}	[33]
k_{cat}	CAD4	KAld	1.33 s^{-1}	[33]
k_{cat}	CAD4	CAld	2.10 s^{-1}	[33]
k_{cat}	CAD4	5CAld	2.52 s^{-1}	[33]
k_{cat}	CAD4	SAld	0.72 s^{-1}	[33]
K_m	CAD4	*p*CAld	47 μM	[33]
K_m	CAD4	KAld	87 μM	[33]
K_m	CAD4	CAld	65 μM	[33]
K_m	CAD4	5CAld	85 μM	[33]
K_m	CAD4	SAld	274 μM	[33]
k_{cat}	CAD3	*p*CAld	1.65 s^{-1}	[33]
k_{cat}	CAD3	KAld	0.80 s^{-1}	[33]
k_{cat}	CAD3	CAld	0.39 s^{-1}	[33]
k_{cat}	CAD3	5CAld	1.47 s^{-1}	[33]

continued on next page

Table C.1.: *continued*

Parameter Type	Enzyme	Metabolite	Value	Reference
k_{cat}	CAD3	SAld	0.74 s^{-1}	[33]
K_m	CAD3	*p*CAld	292 μM	[33]
K_m	CAD3	KAld	581 μM	[33]
K_m	CAD3	CAld	362 μM	[33]
K_m	CAD3	5CAld	534 μM	[33]
K_m	CAD3	SAld	629 μM	[33]
V_{max}	ALDH	*p*CAld	310 pkat/mg protein	[38]
V_{max}	ALDH	CAld	560 pkat/mg protein	[38]
V_{max}	ALDH	SAld	1700 pkat/mg protein	[38]
K_m	ALDH	*p*CAld	19 μM	[38]
K_m	ALDH	CAld	43 μM	[38]
K_m	ALDH	SAld	150 μM	[38]

[a]Estimated by matching measurements of 4CL2 V_{max} in [29] and [236];

Table C.2.: List of parameters estimated by the hierarchical model.

Parameter identity	Reference value[a]	Estimated value[b]	Unit
α_{C4H}	-	5.29±0.52	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{4CL}	-	76.10±7.52	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{CSE}	-	80.07±7.99	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{HCT}	-	7.92±0.80	nmol g FW^{-1} min^{-1} FPKM^{-1}
$\alpha_{C3'H}$	-	94.22±9.36	nmol g FW^{-1} min^{-1} FPKM^{-1}
$\alpha_{CCoAOMT}$	-	52.47±5.25	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{CCR}	-	40.58±4.06	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{COMT}	-	12.85±1.26	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{F5H}	-	0.44±0.04	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{CAD}	-	99.99±9.94	nmol g FW^{-1} min^{-1} FPKM^{-1}
α_{ALDH}	-	0.02±0.01	nmol g FW^{-1} min^{-1} FPKM^{-1}
$[CoA]_{t0}$	9.65±2.81[c]	0.007±0.001	nmol g FW^{-1}
$[SK]_{t0}$	6.22±1.24[d]	1.13±0.11	nmol g FW^{-1}
$[Phe]_{t0}$	26.22±1.24	21.27±2.11	nmol g FW^{-1}
$\frac{k^{KASK}_{cat,\ HCT}}{k^{pCACoA}_{cat,\ HCT}}$	-	56.35±5.57	Unitless
$K^{HCT}_{m,\ pCACoA}$	-	0.012±0.002	nmol g FW^{-1}
$K^{HCT}_{m,\ pCASK}$	-	0.29±0.04	nmol g FW^{-1}
$K^{HCT}_{m,\ KASK}$	-	0.002±0.000	nmol g FW^{-1}
$K^{HCT}_{m,\ KACoA}$	-	0.44±0.06	nmol g FW^{-1}
$K^{HCT}_{m,\ CoA}$	-	0.76±0.08	nmol g FW^{-1}
$K^{HCT}_{m,\ SK}$	-	0.0003±0.0000	nmol g FW^{-1}
$K^{4CL}_{m,\ CoA}$	-	0.002±0.000	nmol g FW^{-1}
$K^{CCoAOMT}_{m,\ KACoA}$	-	0.005±0.001	nmol g FW^{-1}
Influx	3.44	3.44±0.34	nmol g FW^{-1} min^{-1}
K_{lignin}	-	1.85±0.18	min^{-1}
K_{acid}	-	0.067±0.007	min^{-1}

continued on next page

Table C.2.: *continued*

Parameter identity	Reference value	Estimated value	Unit
K_{KFB}	-	0.071±0.007	nmol g FW^{-1} min^{-1} $FPKM^{-1}$

[a]Reference values obtained from Wang et al., (2019);

[b]Estimated with 1,000 MCMC samples;

[c]Summing from *p*CACoA, KACoA, FACoA and free CoA pools;

[d]Summing from *p*CASK, KASK and free SK pools.

D. SUPPLEMENTARY MATERIALS FOR COMBINING GENE REGULATORY NETWORK AND KINETIC MODELING OF LIGNIN BIOSYNTHESIS IN ARABIDOPSIS

D.1 Mass balances for the kinetic model of lignin biosynthesis

$$\frac{d[Phe_{plastid}]}{dt} = v_{ADT} - v_{PCAT} \tag{D.1}$$

$$\frac{d[Phe_{cytosol}]}{dt} = v_{PCAT} - v_{PAL1} - v_{PAL2} - v_{PAL4} \tag{D.2}$$

$$\frac{d[CA]}{dt} = v_{PAL1} + v_{PAL2} + v_{PAL4} - v_{C4H} - v_{acid_sink,\ CA} \tag{D.3}$$

$$\frac{d[pCA]}{dt} = v_{C4H} + v_{ALDH,\ pCAld} - v_{4CL1,\ pCA} - v_{4CL2,\ pCA} - v_{4CL4,\ pCA} - v_{acid\_sink,\ pCA} \tag{D.4}$$

$$\frac{d[KA]}{dt} = v_{CSE} - v_{4CL1,\ KA} - v_{4CL2,\ KA} - v_{4CL4,\ KA} - v_{COMT,\ KA} - v_{acid\_sink,\ KA} \tag{D.5}$$

$$\frac{d[FA]}{dt} = v_{COMT,\ KA} + v_{ALDH,\ CAld} - v_{4CL1,\ FA} - v_{4CL4,\ FA} - v_{F5H,\ FA} - v_{acid\_sink,\ FA} \tag{D.6}$$

$$\frac{d[5FA]}{dt} = v_{F5H,\ FA} - v_{COMT,\ 5FA} - v_{acid\_sink,\ 5FA} \tag{D.7}$$

$$\frac{d[SA]}{dt} = v_{COMT,\ 5FA} + v_{ALDH,\ SAld} - v_{4CL4,\ SA} - v_{acid\_sink,\ SA} \tag{D.8}$$

$$\frac{d[pCASK]}{dt} = v_{HCT,\ pCACoA} - v_{C3'H} \tag{D.9}$$

$$\frac{d[KASK]}{dt} = v_{C3'H} - v_{HCT,\ KASK} - v_{CSE} \tag{D.10}$$

$$\frac{d[SK]}{dt} = v_{HCT,\ KASK} + v_{CSE,\ KASK} - v_{HCT,\ pCACoA} \tag{D.11}$$

$$\frac{d[pCACoA]}{dt} = v_{4CL1,\ pCA} + v_{4CL2,\ pCA} + v_{4CL4,\ pCA} - v_{HCT,\ pCACoA} - v_{CCR,\ pCACoA} \tag{D.12}$$

$$\frac{d[KACoA]}{dt} = v_{4CL1,\ KA} + v_{4CL2,\ KA} + v_{4CL4,\ KA} + v_{HCT,\ KASK} - v_{CCoAOMT} - v_{CCR,\ KACoA} \tag{D.13}$$

$$\frac{d[FACoA]}{dt} = v_{4CL1,\ FA} + v_{4CL4,\ FA} + v_{CCoAOMT} - v_{CCR,\ FACoA} \tag{D.14}$$

$$\frac{d[SACoA]}{dt} = v_{4CL4,\ SA} - v_{CCR,\ SACoA} \tag{D.15}$$

$$\frac{d[CoA]}{dt} = v_{HCT,\ pCACoA} + v_{CCR,\ pCACoA} + v_{CCR,\ KACoA} + v_{CCR,\ FACoA} + v_{CCR,\ SACoA} - v_{4CL1,\ pCA} - v_{4CL2,\ pCA} - v_{4CL4,\ pCA} - v_{4CL1,\ KA} - v_{4CL2,\ KA} - v_{4CL4,\ KA} - v_{4CL1,\ FA} - v_{4CL4,\ FA} - v_{4CL4,\ SA} - v_{HCT,\ KASK} \tag{D.16}$$

$$\frac{d[pCAld]}{dt} = v_{CCR,\ pCACoA} - v_{ALDH,\ pCAld} - v_{CAD3,\ pCAld} - v_{CAD4,\ pCAld} - v_{CAD5,\ pCAld} - v_{lignin\_sink,\ pCAld} \tag{D.17}$$

$$\frac{d[KAld]}{dt} = v_{CCR,\ KACoA} - v_{COMT,\ KAld} - v_{CAD3,\ KAld} - v_{CAD4,\ KAld} - v_{CAD5,\ KAld} - v_{lignin\_sink,\ KAld} \tag{D.18}$$

$$\frac{d[CAld]}{dt} = v_{CCR,\ FACoA} + v_{COMT,\ KAld} - v_{ALDH,\ CAld} - v_{CAD3,\ CAld} - v_{CAD4,\ CAld} - v_{CAD5,\ CAld} - v_{lignin\_sink,\ CAld} \tag{D.19}$$

$$\frac{d[5CAld]}{dt} = v_{F5H,\ CAld} - v_{COMT,\ 5CAld} - v_{CAD3,\ 5CAld} - v_{CAD4,\ 5CAld} - v_{CAD5,\ 5CAld} - v_{lignin\_sink,\ 5CAld} \tag{D.20}$$

$$\frac{d[SAld]}{dt} = v_{CCR,\ SACoA} + v_{COMT,\ 5CAld} - v_{ALDH,\ SAld} - v_{CAD3,\ SAld} - v_{CAD4,\ SAld} - v_{CAD5,\ SAld} - v_{lignin\_sink,\ SAld} \tag{D.21}$$

$$\frac{d[pCAlc]}{dt} = v_{CAD3,\ pCAld} + v_{CAD4,\ pCAld} + v_{CAD5,\ pCAld} - v_{lignin\_sink,\ pCAlc} \tag{D.22}$$

$$\frac{d[KAlc]}{dt} = v_{CAD3,\ KAld} + v_{CAD4,\ KAld} + v_{CAD5,\ KAld} - v_{COMT,\ KAlc} - v_{lignin\_sink,\ KAlc} \tag{D.23}$$

$$\frac{d[CAlc]}{dt} = v_{CAD3,\ CAld} + v_{CAD4,\ CAld} + v_{CAD5,\ CAld} + v_{COMT,\ KAlc} - v_{F5H,\ CAlc} - v_{lignin\_sink,\ CAlc} \tag{D.24}$$

$$\frac{d[5CAlc]}{dt} = v_{CAD3,\ 5CAld} + v_{CAD4,\ 5CAld} + v_{CAD5,\ 5CAld} + v_{F5H,\ CAlc} - v_{COMT,\ 5CAlc} - v_{lignin\_sink,\ 5CAlc} \tag{D.25}$$

$$\frac{d[SAlc]}{dt} = v_{CAD3,\ SAld} + v_{CAD4,\ SAld} + v_{CAD5,\ SAld} + v_{COMT,\ 5CAlc} - v_{lignin\_sink,\ SAlc} \tag{D.26}$$

D.2 Rate equations for the kinetic model of lignin biosynthesis

$$v_{ADT} = \frac{V_{ADT}^{apparent}}{1 + \frac{[Phe_{plastid}]}{K_{i,\ Phe}^{ADT}}} \tag{D.27}$$

$$v_{PCAT} = V_{max,\ PCAT} \frac{[Phe_{plastid}]}{K_{m,\ Phe}^{PCAT}(1 + \frac{[Phe_{cytosol}]}{K_{i,\ Phe}^{PCAT}}) + [Phe_{plastid}]} \tag{D.28}$$

$$v_{PAL1} = V_{max,\ PAL1} \frac{[Phe_{cytosol}]}{K_{m,\ Phe}^{PAL1} + [Phe_{cytosol}]} \tag{D.29}$$

$$v_{PAL2} = V_{max,\ PAL2} \frac{[Phe_{cytosol}]}{K_{m,\ Phe}^{PAL2} + [Phe_{cytosol}]} \tag{D.30}$$

$$v_{PAL4} = V_{max,\ PAL4} \frac{[Phe_{cytosol}]}{K_{m,\ Phe}^{PAL4} + [Phe_{cytosol}]} \tag{D.31}$$

$$v_{C4H} = V_{max,\ C4H} \frac{[CA]}{K_{m,\ CA}^{C4H} + [CA]} \tag{D.32}$$

$$v_{4CL1,\ pCA} = V_{max,\ 4CL1}^{pCA} \frac{\frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL1} K_{m,\ CoA}^{4CL}}}{1 + \frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL1} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL1}} + \frac{[KA]}{K_{m,\ KA}^{4CL1}} + \frac{[FA]}{K_{m,\ FA}^{4CL1}}} \tag{D.33}$$

$$v_{4CL1,\ KA} = V_{max,\ 4CL1}^{KA} \frac{\frac{[KA][CoA]}{K_{m,\ KA}^{4CL1} K_{m,\ CoA}^{4CL}}}{1 + \frac{[KA][CoA]}{K_{m,\ KA}^{4CL1} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL1}} + \frac{[KA]}{K_{m,\ KA}^{4CL1}} + \frac{[FA]}{K_{m,\ FA}^{4CL1}}} \tag{D.34}$$

$$v_{4CL1,\ FA} = V_{max,\ 4CL1}^{FA} \frac{\frac{[FA][CoA]}{K_{m,\ FA}^{4CL1} K_{m,\ CoA}^{4CL}}}{1 + \frac{[FA][CoA]}{K_{m,\ FA}^{4CL1} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL1}} + \frac{[KA]}{K_{m,\ KA}^{4CL1}} + \frac{[FA]}{K_{m,\ FA}^{4CL1}}} \tag{D.35}$$

$$v_{4CL2,\ pCA} = V_{max,\ 4CL2}^{pCA} \frac{\frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL2} K_{m,\ CoA}^{4CL}}}{1 + \frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL2} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL2}} + \frac{[KA]}{K_{m,\ KA}^{4CL2}}} \tag{D.36}$$

$$v_{4CL2,KA} = V_{max,\ 4CL2}^{KA} \frac{\frac{[KA][CoA]}{K_{m,\ KA}^{4CL2} K_{m,\ CoA}^{4CL}}}{1 + \frac{[KA][CoA]}{K_{m,\ KA}^{4CL2} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL2}} + \frac{[KA]}{K_{m,\ KA}^{4CL2}}} \tag{D.37}$$

$$v_{4CL4,\ pCA} = V_{max,\ 4CL4}^{pCA} \frac{\frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL4} K_{m,\ CoA}^{4CL}}}{1 + \frac{[pCA][CoA]}{K_{m,\ pCA}^{4CL4} K_{m,\ CoA}^{4CL}} + \frac{[pCA]}{K_{m,\ pCA}^{4CL4}} + \frac{[KA]}{K_{m,\ KA}^{4CL4}} + \frac{[FA]}{K_{m,\ FA}^{4CL4}} + \frac{[SA]}{K_{m,\ SA}^{4CL4}}} \tag{D.38}$$

$$v_{4CL4,\ KA} = V^{KA}_{max,\ 4CL4} \frac{\frac{[KA][CoA]}{K^{4CL4}_{m,\ KA} K^{4CL}_{m,\ CoA}}}{1 + \frac{[KA][CoA]}{K^{4CL4}_{m,\ KA} K^{4CL}_{m,\ CoA}} + \frac{[pCA]}{K^{4CL4}_{m,\ pCA}} + \frac{[KA]}{K^{4CL4}_{m,\ KA}} + \frac{[FA]}{K^{4CL4}_{m,\ FA}} + \frac{[SA]}{K^{4CL4}_{m,\ SA}}} \tag{D.39}$$

$$v_{4CL4,\ FA} = V^{FA}_{max,\ 4CL4} \frac{\frac{[FA][CoA]}{K^{4CL4}_{m,\ FA} K^{4CL}_{m,\ CoA}}}{1 + \frac{[FA][CoA]}{K^{4CL4}_{m,\ FA} K^{4CL}_{m,\ CoA}} + \frac{[pCA]}{K^{4CL4}_{m,\ pCA}} + \frac{[KA]}{K^{4CL4}_{m,\ KA}} + \frac{[FA]}{K^{4CL4}_{m,\ FA}} + \frac{[SA]}{K^{4CL4}_{m,\ SA}}} \tag{D.40}$$

$$v_{4CL4,\ SA} = V^{SA}_{max,\ 4CL4} \frac{\frac{[SA][CoA]}{K^{4CL4}_{m,\ SA} K^{4CL}_{m,\ CoA}}}{1 + \frac{[SA][CoA]}{K^{4CL4}_{m,\ SA} K^{4CL}_{m,\ CoA}} + \frac{[pCA]}{K^{4CL4}_{m,\ pCA}} + \frac{[KA]}{K^{4CL4}_{m,\ KA}} + \frac{[FA]}{K^{4CL4}_{m,\ FA}} + \frac{[SA]}{K^{4CL4}_{m,\ SA}}} \tag{D.41}$$

$$v_{HCT,\ pCACoA} = V^{pCACoA}_{max,\ HCT} \frac{\frac{[pCACoA][SK]}{K^{HCT}_{m,\ pCACoA} K^{HCT}_{m,\ SK}} - \frac{k^{pCASK}_{cat,\ HCT}[pCASK][CoA]}{k^{pCACoA}_{cat,\ HCT} K^{HCT}_{m,\ pCASK} K^{HCT}_{m,\ CoA}}}{1 + \frac{[pCACoA]}{K^{HCT}_{m,\ pCACoA}} + \frac{[pCASK]}{K^{HCT}_{m,\ pCASK}} + \frac{[KACoA]}{K^{HCT}_{m,\ KACoA}} + \frac{[KASK]}{K^{HCT}_{m,\ KASK}} + \frac{[pCACoA][SK]}{K^{HCT}_{m,\ pCACoA} K^{HCT}_{m,\ SK}} + \frac{[pCASK][CoA]}{K^{HCT}_{m,\ pCASK} K^{HCT}_{m,\ CoA}} + \frac{[KACoA][SK]}{K^{HCT}_{m,\ KACoA} K^{HCT}_{m,\ SK}} + \frac{[KASK][CoA]}{K^{HCT}_{m,\ KASK} K^{HCT}_{m,\ CoA}}} \tag{D.42}$$

$$v_{HCT,\ KASK} = V^{pCACoA}_{max,\ HCT} \frac{\frac{[KASK][CoA]}{K^{HCT}_{m,\ KASK} K^{HCT}_{m,\ CoA}} - \frac{k^{KACoA}_{cat,\ HCT}[KACoA][SK]}{k^{KASK}_{cat,\ HCT} K^{HCT}_{m,\ KACoA} K^{HCT}_{m,\ SK}}}{1 + \frac{[pCACoA]}{K^{HCT}_{m,\ pCACoA}} + \frac{[pCASK]}{K^{HCT}_{m,\ pCASK}} + \frac{[KACoA]}{K^{HCT}_{m,\ KACoA}} + \frac{[KASK]}{K^{HCT}_{m,\ KASK}} + \frac{[pCACoA][SK]}{K^{HCT}_{m,\ pCACoA} K^{HCT}_{m,\ SK}} + \frac{[pCASK][CoA]}{K^{HCT}_{m,\ pCASK} K^{HCT}_{m,\ CoA}} + \frac{[KACoA][SK]}{K^{HCT}_{m,\ KACoA} K^{HCT}_{m,\ SK}} + \frac{[KASK][CoA]}{K^{HCT}_{m,\ KASK} K^{HCT}_{m,\ CoA}}} \tag{D.43}$$

$$v_{C3'H} = V_{max,\ C3'H} \frac{[pCASK]}{K^{C3'H}_{m,\ pCASK} + [pCASK]} \tag{D.44}$$

$$v_{CSE} = V_{max,\ CSE} \frac{[KASK]}{K^{CSE}_{m,\ KASK} + [KASK]} \tag{D.45}$$

$$v_{CCoAOMT} = V_{max,\ CCoAOMT} \frac{[KACoA]}{K^{CCoAOMT}_{m,\ KACoA} + [KACoA]} \tag{D.46}$$

$$v_{COMT,\ KA} = V^{KA}_{max,\ COMT} \frac{\frac{[KA]}{K^{COMT}_{m,\ KA}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.47}$$

$$v_{COMT,\ 5FA} = V^{5FA}_{max,\ COMT} \frac{\frac{[5FA]}{K^{COMT}_{m,\ 5FA}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.48}$$

$$v_{COMT,\ KAld} = V^{KAld}_{max,\ COMT} \frac{\frac{[KAld]}{K^{COMT}_{m,\ KAld}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.49}$$

$$v_{COMT,\ 5CAld} = V^{5CAld}_{max,\ COMT} \frac{\frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.50}$$

$$v_{COMT,\ KAlc} = V^{KAlc}_{max,\ COMT} \frac{\frac{[KAlc]}{K^{COMT}_{m,\ KAlc}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.51}$$

$$v_{COMT,\ 5CAlc} = V^{5CAlc}_{max,\ COMT} \frac{\frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}}{1 + \frac{[KA]}{K^{COMT}_{m,\ KA}} + \frac{[5FA]}{K^{COMT}_{m,\ 5FA}} + \frac{[KAld]}{K^{COMT}_{m,\ KAld}} + \frac{[5CAld]}{K^{COMT}_{m,\ 5CAld}} + \frac{[KAlc]}{K^{COMT}_{m,\ KAlc}} + \frac{[5CAlc]}{K^{COMT}_{m,\ 5CAlc}}} \tag{D.52}$$

$$v_{CCR,\ pCACoA} = V^{pCACoA}_{max,\ CCR} \frac{\frac{[pCACoA]}{K^{CCR}_{m,\ pCACoA}}}{1 + \frac{[pCACoA]}{K^{CCR}_{m,\ pCACoA}} + \frac{[KACoA]}{K^{CCR}_{m,\ KACoA}} + \frac{[FACoA]}{K^{CCR}_{m,\ FACoA}} + \frac{[SACoA]}{K^{CCR}_{m,\ SACoA}}} \tag{D.53}$$

$$v_{CCR,\ KACoA} = V^{KACoA}_{max,\ CCR} \frac{\frac{[KACoA]}{K^{CCR}_{m,\ KACoA}}}{1 + \frac{[pCACoA]}{K^{CCR}_{m,\ pCACoA}} + \frac{[KACoA]}{K^{CCR}_{m,\ KACoA}} + \frac{[FACoA]}{K^{CCR}_{m,\ FACoA}} + \frac{[SACoA]}{K^{CCR}_{m,\ SACoA}}} \tag{D.54}$$

$$v_{CCR,\ FACoA} = V^{FACoA}_{max,\ CCR} \frac{\frac{[FACoA]}{K^{CCR}_{m,\ FACoA}}}{1 + \frac{[pCACoA]}{K^{CCR}_{m,\ pCACoA}} + \frac{[KACoA]}{K^{CCR}_{m,\ KACoA}} + \frac{[FACoA]}{K^{CCR}_{m,\ FACoA}} + \frac{[SACoA]}{K^{CCR}_{m,\ SACoA}}} \tag{D.55}$$

$$v_{CCR,\ SACoA} = V^{SACoA}_{max,\ CCR} \frac{\frac{[SACoA]}{K^{CCR}_{m,\ SACoA}}}{1 + \frac{[pCACoA]}{K^{CCR}_{m,\ pCACoA}} + \frac{[KACoA]}{K^{CCR}_{m,\ KACoA}} + \frac{[FACoA]}{K^{CCR}_{m,\ FACoA}} + \frac{[SACoA]}{K^{CCR}_{m,\ SACoA}}} \tag{D.56}$$

$$v_{ALDH,\ pCAld} = V^{pCAld}_{max,\ ALDH} \frac{\frac{[pCAld]}{K^{ALDH}_{m,\ pCAld}}}{1 + \frac{[pCAld]}{K^{ALDH}_{m,\ pCAld}} + \frac{[CAld]}{K^{ALDH}_{m,\ CAld}} + \frac{[SAld]}{K^{ALDH}_{m,\ SAld}}} \tag{D.57}$$

$$v_{ALDH,\ CAld} = V^{CAld}_{max,\ ALDH} \frac{\frac{[CAld]}{K^{ALDH}_{m,\ CAld}}}{1 + \frac{[pCAld]}{K^{ALDH}_{m,\ pCAld}} + \frac{[CAld]}{K^{ALDH}_{m,\ CAld}} + \frac{[SAld]}{K^{ALDH}_{m,\ SAld}}} \tag{D.58}$$

$$v_{ALDH,\ SAld} = V^{SAld}_{max,\ ALDH} \frac{\frac{[SAld]}{K^{ALDH}_{m,\ SAld}}}{1 + \frac{[pCAld]}{K^{ALDH}_{m,\ pCAld}} + \frac{[CAld]}{K^{ALDH}_{m,\ CAld}} + \frac{[SAld]}{K^{ALDH}_{m,\ SAld}}} \tag{D.59}$$

$$v_{F5H,\ FA} = V^{FA}_{max,\ F5H} \frac{\frac{[FA]}{K^{F5H}_{m,\ FA}}}{1 + \frac{[FA]}{K^{F5H}_{m,\ FA}} + \frac{[CAld]}{K^{F5H}_{m,\ CAld}} + \frac{[CAlc]}{K^{F5H}_{m,\ CAlc}}} \tag{D.60}$$

$$v_{F5H,\ CAld} = V^{CAld}_{max,\ F5H} \frac{\frac{[CAld]}{K^{F5H}_{m,\ CAld}}}{1 + \frac{[FA]}{K^{F5H}_{m,\ FA}} + \frac{[CAld]}{K^{F5H}_{m,\ CAld}} + \frac{[CAlc]}{K^{F5H}_{m,\ CAlc}}} \tag{D.61}$$

$$v_{F5H,\ CAlc} = V^{CAlc}_{max,\ F5H} \frac{\frac{[CAlc]}{K^{F5H}_{m,\ CAlc}}}{1 + \frac{[FA]}{K^{F5H}_{m,\ FA}} + \frac{[CAld]}{K^{F5H}_{m,\ CAld}} + \frac{[CAlc]}{K^{F5H}_{m,\ CAlc}}} \tag{D.62}$$

$$v_{CAD3,\ pCAld} = V^{pCAld}_{max,\ CAD3} \frac{\frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}}}{1 + \frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD3}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD3}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD3}_{m,\ SAld}}} \tag{D.63}$$

$$v_{CAD3,\ KAld} = V^{KAld}_{max,\ CAD3} \frac{\frac{[KAld]}{K^{CAD3}_{m,\ KAld}}}{1 + \frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD3}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD3}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD3}_{m,\ SAld}}} \tag{D.64}$$

$$v_{CAD3,\ CAld} = V^{CAld}_{max,\ CAD3} \frac{\frac{[CAld]}{K^{CAD3}_{m,\ CAld}}}{1 + \frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD3}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD3}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD3}_{m,\ SAld}}} \tag{D.65}$$

$$v_{CAD3,\ 5CAld} = V^{5CAld}_{max,\ CAD3} \frac{\frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}}}{1 + \frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD3}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD3}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD3}_{m,\ SAld}}} \tag{D.66}$$

$$v_{CAD3,\ SAld} = V^{SAld}_{max,\ CAD3} \frac{\frac{[SAld]}{K^{CAD3}_{m,\ SAld}}}{1 + \frac{[pCAld]}{K^{CAD3}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD3}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD3}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD3}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD3}_{m,\ SAld}}} \tag{D.67}$$

$$v_{CAD4,\ pCAld} = V^{pCAld}_{max,\ CAD4} \frac{\frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}}}{1 + \frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD4}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD4}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD4}_{m,\ SAld}}} \tag{D.68}$$

$$v_{CAD4,\ KAld} = V^{KAld}_{max,\ CAD4} \frac{\frac{[KAld]}{K^{CAD4}_{m,\ KAld}}}{1 + \frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD4}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD4}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD4}_{m,\ SAld}}} \tag{D.69}$$

$$v_{CAD4,\ CAld} = V^{CAld}_{max,\ CAD4} \frac{\frac{[CAld]}{K^{CAD4}_{m,\ CAld}}}{1 + \frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD4}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD4}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD4}_{m,\ SAld}}} \tag{D.70}$$

$$v_{CAD4,\ 5CAld} = V^{5CAld}_{max,\ CAD4} \frac{\frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}}}{1 + \frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD4}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD4}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD4}_{m,\ SAld}}} \tag{D.71}$$

$$v_{CAD4,\ SAld} = V^{SAld}_{max,\ CAD4} \frac{\frac{[SAld]}{K^{CAD4}_{m,\ SAld}}}{1 + \frac{[pCAld]}{K^{CAD4}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD4}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD4}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD4}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD4}_{m,\ SAld}}} \tag{D.72}$$

$$v_{CAD5,\ pCAld} = V^{pCAld}_{max,\ CAD5} \frac{\frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}}}{1 + \frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD5}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD5}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD5}_{m,\ SAld}}} \tag{D.73}$$

$$v_{CAD5,\ KAld} = V^{KAld}_{max,\ CAD5} \frac{\frac{[KAld]}{K^{CAD5}_{m,\ KAld}}}{1 + \frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD5}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD5}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD5}_{m,\ SAld}}} \tag{D.74}$$

$$v_{CAD5,\ CAld} = V^{CAld}_{max,\ CAD5} \frac{\frac{[CAld]}{K^{CAD5}_{m,\ CAld}}}{1 + \frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD5}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD5}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD5}_{m,\ SAld}}} \tag{D.75}$$

$$v_{CAD5,\ 5CAld} = V^{5CAld}_{max,\ CAD5} \frac{\frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}}}{1 + \frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD5}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD5}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD5}_{m,\ SAld}}} \tag{D.76}$$

$$v_{CAD5,\ SAld} = V^{SAld}_{max,\ CAD5} \frac{\frac{[SAld]}{K^{CAD5}_{m,\ SAld}}}{1 + \frac{[pCAld]}{K^{CAD5}_{m,\ pCAld}} + \frac{[KAld]}{K^{CAD5}_{m,\ KAld}} + \frac{[CAld]}{K^{CAD5}_{m,\ CAld}} + \frac{[5CAld]}{K^{CAD5}_{m,\ 5CAld}} + \frac{[SAld]}{K^{CAD5}_{m,\ SAld}}} \tag{D.77}$$

$$v_{acid\_sink,\ CA} = K_{acid}[CA] \tag{D.78}$$

$$v_{acid\_sink,\ pCA} = K_{acid}[pCA] \tag{D.79}$$

$$v_{acid\_sink,\ KA} = K_{acid}[KA] \tag{D.80}$$

$$v_{acid\_sink,\ FA} = K_{acid}[FA] \tag{D.81}$$

$$v_{acid\_sink,\ 5FA} = K_{acid}[5FA] \tag{D.82}$$

$$v_{acid\_sink,\ SA} = K_{acid}[SA] \tag{D.83}$$

$$v_{lignin\_sink,\ pCAld} = K_{lignin}[pCAld] \tag{D.84}$$

$$v_{lignin\_sink,\ KAld} = K_{lignin}[KAld] \tag{D.85}$$

$$v_{lignin\_sink,\ CAld} = K_{lignin}[CAld] \tag{D.86}$$

$$v_{lignin\_sink,\ 5CAld} = K_{lignin}[5CAld] \tag{D.87}$$

$$v_{lignin\_sink,\ SAld} = K_{lignin}[SAld] \tag{D.88}$$

$$v_{lignin\_sink,\ pCAlc} = K_{lignin}[pCAlc] \tag{D.89}$$

$$v_{lignin\_sink,\ KAlc} = K_{lignin}[KAlc] \tag{D.90}$$

$$v_{lignin\_sink,\ CAlc} = K_{lignin}[CAlc] \tag{D.91}$$

$$v_{lignin\_sink,\ 5CAlc} = K_{lignin}[5CAlc] \tag{D.92}$$

$$v_{lignin\_sink,\ SAlc} = K_{lignin}[SAlc] \tag{D.93}$$