ABSTRACT

Title of Dissertation:	ANALYSIS OF OBJECTIVES AND CONSTRAINTS TOWARDS PREDICTIVE MODELING OF COMPLEX METABOLISM
	Navadeep Boruah, Doctor of Philosophy, 2020
Dissertation directed by:	Dr. Ganesh Sriram, Associate Professor, Dept. of Chemical and Biomolecular Engineering, Committee Chair

The central theme of this dissertation is predictive modeling of metabolism in complex biological systems with genome-scale stoichiometric metabolic models to (i) gain nontrivial insight on cellular metabolism and (ii) provide justifications for hitherto unexplained metabolic phenomena. The crux of high-quality predictive modeling with genome-scale stoichiometric metabolic models is appropriate selection of (i) a biologically relevant objective function and (ii) a set of constraints based on experimental data. However, in many complex systems, like a plant tissue with its wide array of specialized cells, a biological objective is not always apparent. Additionally, generation of experimental data to develop biochemically relevant constraints can be nearly impossible in systems that cannot be cultured under a controlled environment for the duration of an experiment. Such limitations necessitate careful reformulation of the biological question, development of novel methods and data analysis strategies.

Here, we push the boundaries of predictive modeling by demonstrating its first application in deciphering hitherto unexplained metabolic phenomena and in developing novel hypotheses on metabolism. Towards achieving this goal, we developed several novel approaches and employed them in diverse biological systems. Firstly, we investigated the selection of carbohydrate degrading pathway employed by *Saccharophagus degradans*, an aerobic cellulosic marine bacterium. Flux balance analyses of its growth in nutrient rich hypoxic marine environment predicted that the selection of carbohydrate degrading pathway is possibly influenced by inorganic nutrient availability. Secondly, multi-tissue genome-scale metabolic modeling of *Populus trichoparpa*, a perennial woody tree, and analyses with a novel strategy based on multiple biologically relevant metrics provided a metabolic justification for the predominance of glutamine as the predominant nitrogen transport amino acid for internal nitrogen recycling. Thirdly, predictive modeling of maize grain filling predicted amino acid fermentation as a mechanism for expending excess reductant cofactors for continual starch synthesis in the hypoxic environment of endosperm. Finally, we developed bilevel optimization framework to integrate publicly available transcriptome datasets with metabolic networks. This framework predicted accumulation of specific classes of maize endosperm storage protein at distinct stages of grain filling. We anticipate that the employment of these aforementioned approaches in other biological systems will lead to the generation of a wide array of nontrivial hypothesis on cellular metabolism and to develop targeted experiments to validate them.

ANALYSIS OF OBJECTIVES AND CONSTRAINTS TOWARDS PREDICTIVE MODELING OF COMPLEX METABOLISM

by

Navadeep Boruah

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Dr. Ganesh Sriram, Associate Professor, Dept. of Chemical and Biomolecular Engineering, Chair
Dr. Panagiotis Dimitrakopoulos, Associate Professor, Dept. of Chemical and Biomolecular Engineering
Dr. Jeffery B. Klauda, Associate Professor, Dept. of Chemical and Biomolecular Engineering
Dr. Zhongchi Liu, Professor, Dept. of Cell Biology and Molecular Genetics
Dr. Amy J. Karlsson, Associate Professor, Dept. of Chemical and Biomolecular Engineering

© Copyright by
Navadeep Boruah
2020

Dedication

I dedicate this dissertation to my wife Michelle Boruah for her unwavering support throughout the duration of PhD program and beyond. Without her constant encouragement and support I would have never finished this dissertation.

Acknowledgements

Firstly, I would like to thank my advisor, Dr. Ganesh Sriram, for accepting me in his research group and introducing me to the world of biological sciences.

Over the last five years or so I had the opportunity to work with some very talented undergraduate researchers in the lab. In particular, I am indebted to Brendon Cox for spending a massive amount of time curating the maize model and performing the simulations. Additionally, David Samson and Daniel Connelly also provided invaluable help on various projects.

I thank all my past and present co-workers in the lab, especially Xiaofeng Zhang, Daniel Lugar, Shriramprasad Venkatesan Andrew Quinn and Yuting Zheng for all the fruitful discussions in the lab.

I express my gratitude to Dr. Gary Coleman and Dr. Ashish Misra for their insight and help on my research on poplar metabolism. I also thank Dr. Heven Sze for many meaningful discussions over the years.

I am grateful to my Ph.D. proposal and dissertation committee members, Dr. Panagiotis Dimitrakopoulos, Dr. Jeffery Klauda, Dr. Zhongchi Liu and Dr. Amy Karlsson for their feedback on this dissertation.

I thank my parents Chandra Boruah and Bharati Gogoi Boruah for their support in my decision of quitting my job to attend graduate school. Without my wife Michelle Boruah's endless patience and support I would not have been able to finish this dissertation. My brother Nilotpal has been a pillar of strength for me over the last few years. And without Sourabh Bora's persistence and help I probably would have never left the shores of India.

Finally, I acknowledge my funding source for this work, the National Science Foundation (IOS – 0922650, CBET - 1134115, MCB - 1517671).

Table of Contents

Dedication ii
Acknowledgements iii
Table of Contents iv
List of Figures vii
List of Tables ix
List of Abbreviations x
Chapter 1 : Introduction 1
1.1 Cells as Microscopic Factories 1
1.2 Cellular Metabolism and Metabolic Flux 2
1.2.1 Flux Balance Analysis 2
1.2.2 ^{13}C-MFA 4
1.3 Metabolic Flux Analysis in Strain Development 6
1.4 Computational Methods in Predictive Modeling 7
1.5 Challenges in Predicting Plant Metabolism 8
1.6 Outline of Dissertation 9
Chapter 2 : Investigation of Carbohydrate Metabolism of a Heterotrophic Cellulosic Bacterium 11
2.1 Background 11
2.1.1. *Saccharophagus degradans* a model aerobic cellulosic microbe 12
2.2 Materials and Methods 12
2.2.1 Construction of Draft Model 12
2.2.2 Gap Filling 13
2.2.3 Substrate Cycles 14
2.2.4 Generic Reactions and Dead-End Metabolites 14
2.2.5 Mathematical Formulation 14
2.3 Results 14
2.3.1 Metabolism of Glucose, Cellobiose and Xylose by *Sde* 15
2.3.2 Phosphate Limitation Dictate Pathway Selection 20
2.3.2 Dual Analysis Points to Possible Amino Acid Fermentation 21
2.4 Conclusion 21
Chapter 3 : A Genome-Scale Stoichiometric Model of *Populus trichocarpa* for Investigation of Tree Metabolism: Seasonal Nitrogen Mobilization and Isoprene Emission 22
3.1 Background 22
3.2 Construction and curation of PoplarGSM1.0 23
3.2.1 Generation of raw model draft 23
3.2.2 Assigning subcellular compartmentalization to reactions 24
3.2.3 Gap filling 24
3.2.4 Generic reactions and dead-end metabolites 25
3.2.5 Substrate cycles 25
3.2.6 Development of the biomass reaction 25
3.3 Results 26
3.3.1 Nitrogen storage in bark storage protein and its remobilization during spring 27

3.3.2 Emission of isoprene on hot days and intense sunlight 52
3.4 Discussion 55
3.4.1 Efficiency of N mobilization for early spring growth might be a factor in selection of N transport amino acid 55
3.4.2 Multi-tissue GSM and careful selection of constraints is essential for capturing meaningful information on complex metabolic events 56
3.4.3 Interpretation of simulations based on multiple biologically relevant metrics is necessary for understanding metabolism in complex systems 56
3.4.4 Context specific objective function is essential for specialized metabolism in plant tissues 56
3.5 Conclusion 57
Chapter 4 : A Genome-Scale Stoichiometric Model of *Zea mays* for Investigation Endosperm Metabolism: Role of Pyruvate Phosphate Dikinase and Amino Acid Fermentation on Energetics of Grain Filling 58
4.1 Background 58
4.2 Materials and Methods 61
4.2.1 Genetic modification of maize and kernel collection 61
4.2.2 Metabolite extraction and analysis 62
4.2.3 Construction of the draft GSM 62
4.2.4 Gap filling and substrate cycles 64
4.2.5 Generic Reactions and Dead-End Metabolites 64
4.2.6 Development of the biomass reaction 65
4.2.7 Objective functions and analyses 65
4.2.7 Software 65
4.3 Results 65
4.3.1 Knockdown of *ppdk* changes endosperm metabolite profiles 65
4.3.2 FBA of PPDK deficient Tx endosperm predicted altered Gln uptake and amino acid biosynthesis 67
4.3.3 Gln uptake and specific amino acid biosynthesis crucial to maximal starch synthesis under hypoxia 68
4.3.4 PPDK was predicted to provide flexibility of ADP-Glc synthesis across compartments 71
4.3.5 Pro biosynthesis is predicted to be essential for Gln uptake reduction 71
4.4 Discussion 72
4.4.1 An iterative approach to genome-scale FBA enables faster hypothesis development 73
4.4.2 Essentiality of Gln uptake in the simulations is necessitated by reductant cofactor utilization 73
4.4.3 Gln-supported metabolism explains spatiotemporal variations in endosperm metabolite profile 74
4.4.4 PPDK permits redundancy of ADP-Glc synthesis and subsequent PPi recycling in cytosol and plastid 74
4.4.5 Predictive modeling with limited constraints: a best-case scenario in a worst-case scenario 75
4.5 Conclusion 75

Chapter 5 : Integration of transcriptome data with a genome-scale stoichiometric model of maize for investigation of storage protein accumulation in endosperm 76
5.1 Background 76
5.2 Integration of Transcriptome Datasets with GSMs: An Overview of Past Research 78
5.3 Materials and Methods 79
5.3.1 Model Curation 79
5.3.2 Transcriptome Datasets 79
5.3.3 Curation of gene-to-reaction association 80
5.3.4 Challenges, Limitations of Constraints/Datasets and Assumptions 80
5.3.5 Assumptions in Our Framework 81
5.3.6 Logistics Equation 81
5.3.7 Computational Framework 82
5.3.8 Software 85
5.4 Results 85
5.4.1 Single level optimization framework cannot predict zein synthesis 85
5.4.2 Bi-level optimization with maximum nitrogen storage in zeins as the inner objective 86
5.4.3 Bi-level optimization with maximum starch synthesis as the inner objective 87
5.4.4 Bi-level optimization with maximum Suc uptake as the inner objective 88
5.5 Discussion 89
5.6.1 Improved predictive power of bilevel optimization 89
5.6.2 Iterative Approach Not Entirely Eliminated 90
5.6.3 Predictive modeling with transcriptome matching as the inner objective 91
5.6 Conclusion 91
Chapter 6 : Summary and Future Outlook 92
6.1 Summary 92
6.2 Future Outlook 92
References 94

List of Figures

FIGURE 1.1. GRAPHICAL REPRESENTATION OF A SMALL FBA PROBLEM WITH TWO DEGREES OF FREEDOM. 4
FIGURE 1.2. SCHEMATICS OF ^{13}C-MFA RELIES ON UNIQUE ISOTOPIC SIGNATURES OF DIFFERENT PATHWAYS. 5
FIGURE 1.3. OVERVIEW OF COMPUTATIONAL STRAIN DESIGN. 6
FIGURE 2.1. WORKFLOW OF GENOME-SCALE MODEL CONSTRUCTION OF *S. DEGRADANS* STRAIN 2-40. 13
FIGURE 2.2. ALTERED FLUX DISTRIBUTIONS OF *SDE* ON GLC PREDICTED UPON BLOCKAGE OF *PFK*
FIGURE 2.3. ALTERED FLUX DISTRIBUTIONS OF *SDE* ON XYL PREDICTED UPON BLOCKAGE OF *PFK*. 17
FIGURE 2.4. TCA CYCLE FLUXES IN *SDE* VARY DEPENDING UPON EMP VS ED PATHWAY ACTIVITY 18
FIGURE 2.5. FLUX DISTRIBUTIONS OF *SDE* BLOCKED EMP PATHWAY ENZYMES (*PFK*, *GAPDH* OR *PGK*) AND PARTIAL RESTRICTION OF ED PATHWAY AND PPP. 19
FIGURE 2.6. SHADOW PRICES OF INTERNAL AND EXTRACELLULAR REACTIONS UNDER ENZYMATIC LIMITATION AND HYPOXIA 20
FIGURE 3.1. WORKFLOW OF POPLARGSM1.0 RECONSTRUCTION AND ANALYSIS. 26
FIGURE 3.2. ORGANELLAR DISTRIBUTIONS OF UNIQUE METABOLIC REACTIONS AND METABOLITES IN POPLARGSM1.0. 27
FIGURE 3.3. CONCENTRATION AND PERCENTAGE OF AMINO ACIDS IN POPLAR BARK AND XYLEM 28
FIGURE 3.4. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL PREDICTS GLUTAMINE TO BE THE MOST ENERGY- AND NITROGEN-CONSERVING AMINO ACID FOR TRANSPORT OF NITROGEN FROM BARK TO SHOOT IN SPRING, CONSISTENT WITH FIELD OBSERVATIONS 30
FIGURE 3.5. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL, WITH ZERO RUBISCO ACTIVITY IN THE SHOOTS, PREDICTS MULTIPLE AMINO ACIDS TO BE EQUALLY EFFICIENT FOR TRANSPORT OF NITROGEN FROM BARK TO SHOOT IN SPRING, CONSISTENT WITH FIELD OBSERVATIONS. 34
FIGURE 3.6. OPTIMAL ROUTE FOR EARLY SPRING BSP MOBILIZATION AS PREDICTED BY FBA 35
FIGURE 3.7. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL PREDICTS GLUTAMINE TO BE SUBOPTIMAL, WHEN CO_2 UPTAKE AND PHOTOSYNTHESIS ARE PERMITTED, FOR DURING LEAF NITROGEN MOBILIZATION IN FALL. 38
FIGURE 3.8. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL PREDICTS GLUTAMINE TO BE SUBOPTIMAL ON OPTIMAL REDUCTANT UTILIZATION DURING LEAF NITROGEN MOBILIZATION IN FALL, WITH TRANSPORT OF CO_2 POOL RELEASED IN BARK TO SENESCING LEAVES AND PHOTON UPTAKE WITHOUT ADDITIONAL CO_2 UPTAKE IN LEAVES 40
FIGURE 3.9. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL PREDICTS GLUTAMINE TO BE SUBOPTIMAL ON MULTIPLE BIOLOGICALLY RELEVANT METRICS DURING LEAF NITROGEN MOBILIZATION IN FALL, WITH PHOTON UPTAKE BUT WITHOUT ANY CO_2 UPTAKE OR MIXING OF RELEASED CO_2 POOLS 45
FIGURE 3.10. FLUX BALANCE ANALYSIS (FBA) ON A POPLAR GENOME-SCALE METABOLIC MODEL PREDICTS GLUTAMINE TO BE SUBOPTIMAL ON MULTIPLE BIOLOGICALLY RELEVANT METRICS DURING LEAF NITROGEN MOBILIZATION IN FALL, WITHOUT REDUCTANT GENERATED FROM PHOTOSYNTHESIS 47
FIGURE 3.11. OPTIMAL ROUTE FOR LEAF N MOBILIZATION DURING LEAF SENESCENCE AS PREDICTED BY FBA. A, FLUX MAP DISPLAYING THE REACTIONS IN THE LEAF CELL OF POPLAR DURING MOBILIZATION OF LEAF N CONTENT TO ONLY ARG. 51

FIGURE 3.12. CENTRAL CARBON METABOLISM AND ISOPRENE BIOSYNTHESIS PATHWAYS OF POPLAR.....54
FIGURE 4.1. SCHEMATIC REPRESENTATION OF NUTRIENT FLOW IN MAIZE. MAIZE LEAVES PHOTOSYNTHETICALLY FIX CO_2 AND THEN CONVERT THE PHOTOSYNTHATE TO SUCROSE.....59
FIGURE 4.2. WORKFLOW OF *ZEA MAYS* GENOME-SCALE MODEL RECONSTRUCTION ...63
FIGURE 4.3. TOTAL ION COUNTS OF METABOLITES DETECTED IN WT AND TX ENDOSPERMS.....66
FIGURE 4.4. FLUX BALANCE ANALYSIS OF STARCH FILLING IN WT AND TX MAIZE ENDOSPERM. 68
FIGURE 4.5. FLUX VARIABILITY ANALYSIS OF STARCH FILLING IN WT MAIZE ENDOSPERM. 69
FIGURE 4.6. FLUX VARIABILITY ANALYSIS OF STARCH FILLING IN TX MAIZE ENDOSPERM. 70
FIGURE 4.7. ITERATIVE METHODOLOGY EMPLOYED IN RESTRICTING SPECIFIC METABOLITE EXCRETION IN FBA SIMULATIONS. 72
FIGURE 5.1. DISTRIBUTION OF ALL PROTEINOGENIC AMINO ACIDS IN ZEIN PROTEINS. 78
FIGURE 5.2. PLOT OF ESTIMATED PROBABILITIES p_{jk} AGAINST THE ASSUMED PROBABILITIES p_{y} 82
FIGURE 5.3. PREDICTIONS OF ENDOSPERM METABOLISM WITH TRANSCRIPTOME MATCHING AS THE SOLE OBJECTIVE. 86
FIGURE 5.4. PREDICTIONS OF ENDOSPERM METABOLISM WITH A BI-LEVEL OPTIMIZATION FRAMEWORK WITH MAXIMUM STORAGE OF N IN ZEINS AS THE INNER OBJECTIVE 87
FIGURE 5.5. PREDICTIONS OF ENDOSPERM METABOLISM WITH A BI-LEVEL OPTIMIZATION FRAMEWORK WITH MAXIMUM STARCH SYNTHESIS AS THE INNER OBJECTIVE.....88
FIGURE 5.6. PREDICTIONS OF ENDOSPERM METABOLISM WITH A BI-LEVEL OPTIMIZATION FRAMEWORK WITH MAXIMUM SUC UPTAKE AS THE INNER OBJECTIVE.....89
FIGURE 5.7. DISTRIBUTION OF ALL PROTEINOGENIC AMINO ACIDS IN AT THE FIVE DIFFERENT TIME POINTS. 90

List of Tables

TABLE 1. CONSTRAINTS ON MULTI-TISSUE FBA FOR BSP MOBILIZATION IN SPRING 32
TABLE 2. CONSTRAINTS ON SINGLE TISSUE FBA FOR BSP MOBILIZATION IN THE BARK TISSUE SPRING 33
TABLE 3. CONSTRAINTS ON MULTI-TISSUE FBA FOR LEAF N MOBILIZATION IN FALL 42
TABLE 4. CONSTRAINTS ON MULTI-TISSUE FBA FOR LEAF N MOBILIZATION IN FALL 43
TABLE 5. CONSTRAINTS ON MULTI-TISSUE FBA FOR LEAF N MOBILIZATION IN FALL 49
TABLE 6. CONSTRAINTS ON MULTI-TISSUE FBA FOR LEAF N MOBILIZATION IN FALL 50

List of Abbreviations

Abbreviation	Metabolite Name
2-IC	2-ketoisocaproate
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
5,1 THF	5,10-methylene-tetrahydrofolate
6PG	6-phosphogluconate
AA	amino acid
aAAA	α-aminoadipate
Ac	acetate
AcAc	acetoacetate
AcAcCoA	acetoacetyl-CoA
AcCoA	acetyl-CoA
ADP	adenosine diphosphate
ADP-Glc	ADP-glucose
ADSUC	adenyl-succinate
aKB	α-ketobutyrate
aKG	α-ketoglutarate
aKMV	α-keto-β-methyl-valerate
aKV	α-keto-isovalerate
Ala	alanine
AMP	adenosine monophosphate
Arg	arginine
ArgSuc	arginino-succinate
Asn	asparagine
Asp	aspartate
ATP	adenosine-triphosphate
B-Ala	β-alanine
CarP	carbamoyl-phosphate
CH3CHO	acetaldehyde
CIT	citrate

Citru	citrulline
CroCoA	crotonyl-CoA
CTP	cytidine-triphosphate
Dhap	dihydroxyacetone phosphate
DMAPP	dimethylallyl diphosphate
DPG	2,3-bisphosphoglycerate
DXP	deoxyxylulose-5-phosphate
E4P	erythrose-4-phosphate
EtOH	ethanol
F6P	fructose-6-phosaphate
FBP	fructose-1,2-biphosphate
fGlu	formylglutamate
Fru	fructose
FUM	fumarate
G1P	glucose-1-phosphate
G6P	glucose-6-phosaphate
GABA	gamma-aminobutyrate
GAP	glyceraldehyde-3-phosphate
Glc	glucose
Gln	glutamine
Glu	glutamate
Gly	glycine
GOx	glyoxylate
GTP	guanosine-triphosphate
HCHO	formaldehyde
HCO3-	bicarbonate
HMBDP	(E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate
HMGCoA	hydroxymethylglutaryl-CoA
ICit	isocitrate
Ile	isoleucine
IPP	isopentenyl diphosphate

Lac	lactate
Leu	leucine
Lva	levulinic acid
Lys	lysine
MAL	malate
MEP	2-methylerylthritol 4-phosphate
Mgly	methyl-glyoxylate
MSA	malonate semialdehyde
MTHF	methylfolate
MVAPP	mevalonate diphosphate
NH3	ammonia
NH4	ammonium
Nleu	norleucine
NO3-	nitrate
Nval	norvaline
OAA	oxaloacetate
Orn	ornithine
P5C	pyrroline-5-carboxylate
P5P	pentose-5-phosphate
PEP	phosphoenolpyruvate
PhAc	phenylacetate
PPi	pyrophosphate
Pro	proline
ProCoA	propionyl-CoA
PRPP	phosphoribosylpyrophosphate
Pyr	pyruvate
R5P	ribose-5-phosphate
Ri5P	ribulose-5-phosphate
S7P	sedoheptulose-7-phosphate
SCoA	succinyl-CoA
Ser	serine

SUC	succinate
THF	tetrahydrofolate
THP	isocitrate
UBL	ubiquinol
UBN	ubiquinone
UDP-Glc	UDP-glucose
UrC	urocanate
UTP	uridine-triphosphate
Val	valine
X5P	xylulose-5-phosphate
Xyl	xylose

Chapter 1 : Introduction

1.1 Cells as Microscopic Factories

Designer cells with optimized metabolic pathways capable of cost-effective production of desired compounds is of tremendous interest to industrial biotechnology (Quin and Schmidt-Dannert, 2014). Utilization of genetically modified designer microbes producing a range of high value compounds is well established in multiple industries. The fast growth, relative ease of rapid genetic manipulation and the presence of infrastructure required for microbial production have been the primary drivers behind the adoption of microbial cells as factories (Tatsis and O'Connor, 2016). Microbes have a limited inventory of endogenous metabolites. Compared to microorganisms, plants synthesize a huge array of compounds, some of which have been used as dyes, flavors, fragrances, stimulants, hallucinogens, insecticides and therapeutic agents etc. (Aharoni and Galili, 2011). Over the last two decades research on biosynthesis of natural plant products with metabolic engineering has seen a major thrust with some incredible achievements (Lau et al., 2014). The most widely used approach in harnessing this extraordinary biosynthetic capacity of plants is reconstitution of plant metabolic pathways synthesizing a compound of interest in heterologous systems, primarily microorganisms (Lau et al., 2014; O'Connor, 2015). However, successful expression of plant metabolic pathways in heterologous systems requires extensive optimization and is fraught with failure (O'Connor, 2015). Additionally, expression in heterologous hosts expropriates the advantage of plants not requiring an exogenous carbon source. Engineering plants with the objective of enhancing synthesis of valuable metabolites were intractable largely due to the lack of options for elucidation of enzymes and metabolic pathways, until recently (Schilmiller et al., 2012). The advent and progress of systems biology has provided an avenue for rapid elucidation of plant metabolism and its controlling factors. However, a wide range of fundamental questions on plant metabolism need to be answered before achieving the ultimate objective of using plants as bio-factories.

Systems biology employs a data driven approach to discovery compared to traditional method of purification and analysis of individual enzymes (Schilmiller et al., 2012). Often multiple 'omics' datasets are integrated with sophisticated computational frameworks to develop and test hypothesis. Rapid developments in genomics, transcriptomics, proteomics, and metabolomics technologies coupled with decreasing cost of whole genome sequencing (WGS) have enabled the generation of massive datasets for use with systems biology tools. The biological processes governing cell metabolism are integrated through networks of metabolic reactions and constituents (Jeong et al., 2000). Experimental datasets are employed in conjunction with the biological networks to test hypothesis on biological function of a network component, predict dynamics of networks under perturbations and to provide insight into the governing factors (Baghalian et al., 2014; Kumar et al., 2016). Traditional non-targeted methods reliant on random mutations and natural gene transfer are suboptimal in development of engineered strains with desired properties (Baghalian et al., 2014; O'Connor, 2015). Systems biology methods like plant metabolic network analysis address this limitation by providing a rigorous basis for devising metabolic engineering strategies (Baghalian et al., 2014).

1.2 Cellular Metabolism and Metabolic Flux

The goal of computational modeling in biology is to be able to predict phenotypes arising from molecules and their interactions (Karr et al., 2012). Whole-cell models covering the entire genome and capable of accurate phenotype prediction is the holy grail of computational modeling of cell metabolism. However, phenotype prediction with whole-cell computational modeling is in its infancy due to the complexity of the task and the lack of data on the kinetics and regulatory aspects of metabolism (Macklin et al., 2014; Goldberg et al., 2018). The metabolic flux profile of a cell is the ultimate reflection of its phenotype. Consequently, much of the research on predictive modeling of metabolism has focused towards estimation or prediction of metabolic flux profiles in cells. The role of metabolic flux analysis, perhaps the most commonly used systems biology method, has changed from the secondary role of *a posteriori* phenotype profiling to the vital role of guiding desired phenotype engineering.

Metabolic flux profiling in cells can broadly be categorized into *a priori* prediction of metabolic flux and *a posteriori* estimation of metabolic flux with ^{13}C-metabolix flux analysis. *A priori* prediction of metabolic flux employs the reaction network of an organism with experimental data on nutrient uptake and product synthesis with an objective function like maximization of biomass maximization to predict the optimal state of cell metabolism. Stoichiometric modeling methods like flux balance analysis are the most commonly employed predictive modeling techniques. Many of these predictive modeling methods have found routine applications in computation guided microbial strain optimization. On the other hand, *a posteriori* metabolic flux analysis employs feeding substrates labelled with stable isotopes, often ^{13}C or ^{15}N, strategic locations in the substrate molecule. When the substrate is metabolized by an organism, labelled atoms are distributed through the metabolic network to different metabolites. Metabolic pathways with unique patterns of bond formation and breakage leaves unique isotopic signatures in downstream compounds which are then analyzed by sophisticated instrumental methods like gas chromatography mass spectrometry (GC-MC) or nuclear magnetic resonance spectroscopy (NMR). The experimental data is employed with a global optimization method to estimate the metabolic fluxes in the cell. Estimation of metabolic fluxes by employing ^{13}C-isotope labeled substrates is known as ^{13}C-labelled metabolic flux analysis (^{13}C-MFA). We introduce flux balance analysis and ^{13}C-MFA and discuss the strengths and weaknesses of the methods in the context of plant metabolic networks in the following subsections.

1.2.1 Flux Balance Analysis

Flux balance analysis (FBA) is a linear programing (LP) based method for analyzing the flow of metabolites through a metabolic network (Orth et al., 2010). The crux of FBA is a mass balanced reconstruction of a metabolic network in which the reactions in the metabolic network is converted to a set of linear mass balance equations (**Figure 1.1**) (Kauffman et al., 2003; Orth et al., 2010). These mass balance equations (eqn 0.2) along with bounds on flux values (eqn 0.3) including experimentally measured parameters e.g. substrate uptake rates, the growth rate of the organism, excretion rate of metabolites by

cells etc. constitute the constraints of the linear optimization problem of FBA (Kauffman et al., 2003; Orth et al., 2010). A biologically relevant objective function (eqn 0.1) is optimized with these constraints to solve the linear programing based FBA problem resulting in a metabolic flux distribution (Orth et al., 2010). By virtue of being an LP problem, the predicted optimal flux distribution lies on a corner of the permissible solution space (Orth et al., 2010). When the objective function is linear the FBA problem is P-complete. This is significant since even very large metabolic networks consisting of thousands of metabolic reactions can be solved in a very short time. Multiple packages to solve FBA problems have been developed to efficiently solve such problems (Becker et al., 2007; Ebrahim et al., 2013; Lakshmanan et al., 2014). FBA can be expressed as the following linear programing problem:

$$\max \; c_j v_j \qquad \text{(eqn 0.1)}$$

s.t.

$$S_{ij} v_j = \qquad \text{(eqn 0.2)}$$

$$v_{j,min} \leq v_j \leq v_{j,max} \qquad \text{(eqn 0.3)}$$

The FBA problem is the primal linear programing problem. Each primal problem has a dual problem associated with it (Maranas and Zomorrodi, 2016). The dual of the FBA problem can be written as following:

$$\min \; u_j^{ub} v_{j,max} - u_j^{lb} v_j \qquad \text{(eqn 0.4)}$$

s.t.

$$S_{ji} w_i + u_j^{ub} - u_j^{lb} = c_j \qquad \text{(eqn 0.5)}$$

$$w_i, u_j^{ub}, u_j^{lb} \geq 0 \qquad \text{(eqn 0.6)}$$

The variables w_i, u_j^{ub}, u_j^{lb} are the dual variables associated with the equations (eqn 0.2) and (eqn 0.3) of the primal FBA problem. At optimality the values of the primal and dual objectives are equal. The dual variables, also called shadow prices, indicate the change in optimal value of the primal objective function if the constraint associated with the dual variable is relaxed by one unit.

The major drawback of FBA is possibility of non-uniqueness of the predicted solutions as there exists a large degree of freedom due to the number of reaction fluxes being much larger than the number of mass balance equations. Presence of large number of reversible reactions in the metabolic network often results in multiple pathways that could be equally optimal in fulfilling an objective (Orth et al., 2010). In the presence of parallel equally optimal pathways, the branching of fluxes through the pathways cannot be determined (Wiechert, 2001). Additionally, network reconstruction for FBA requires significant amount of scrutiny. The presence of even a single unbalanced reaction in the network could lead to a futile cycle, a result of two metabolic pathways operating in opposite directions, which could lead to unlimited generation or consumption of mass violating the 1st law of thermodynamics. If the directionalities of reactions are not carefully assigned it could potentially lead to a substrate cycle with unlimited generation or conservation of energy (ATP or reductant). Additionally, the selection of the objective function, which significantly impacts the predictions, in a nontrivial task.

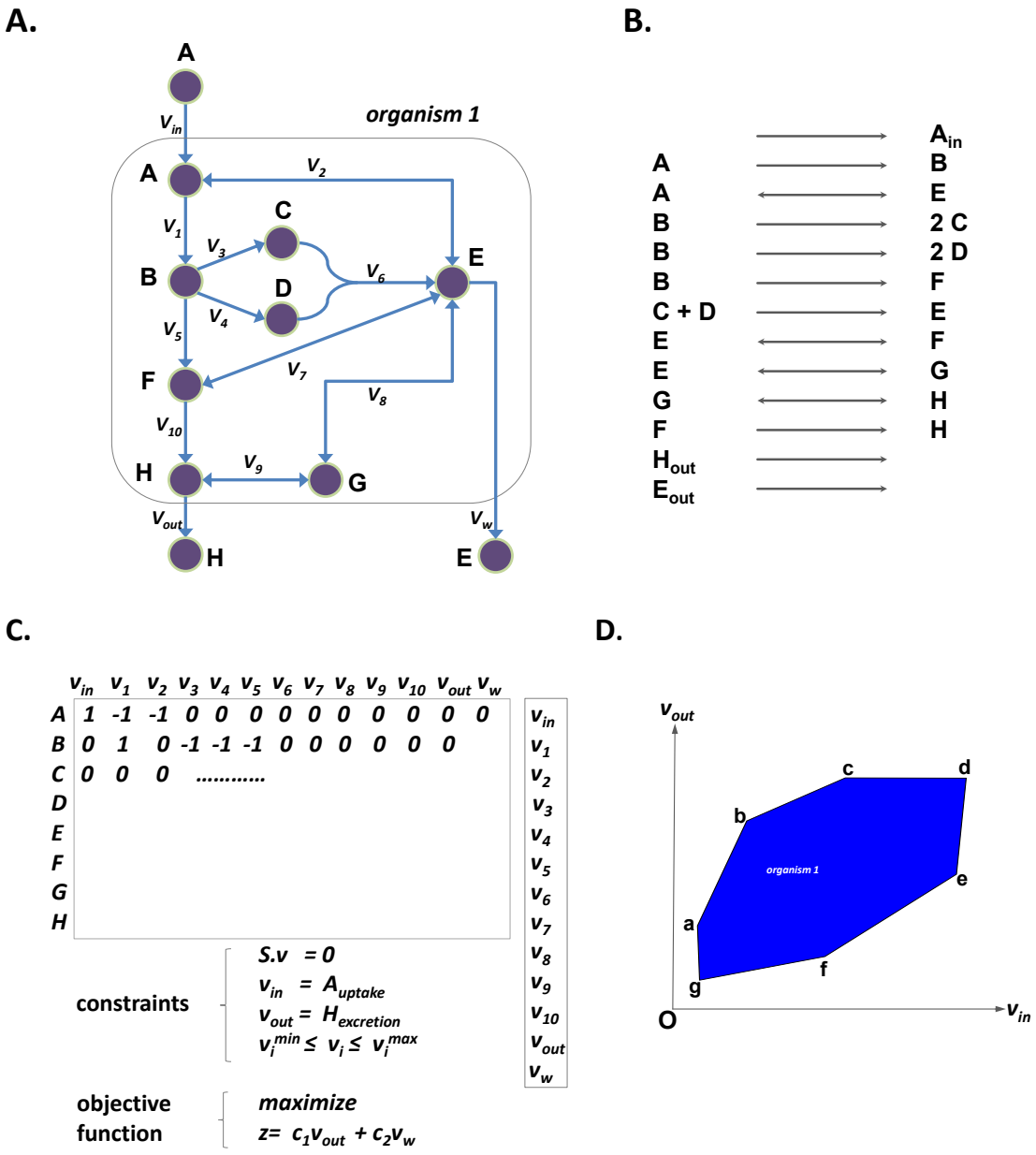


Figure 1.1. Graphical representation of a small FBA problem with two degrees of freedom. A, a reconstructed metabolic network in of an organism with 13 reactions and 11 metabolites; B, list of reactions in the metabolic network; C, the set of reactions in the network converted to a set of linear mass balance equations, with a set of constraints on the flux values and a biological objective; D, two-dimensional feasible solution space or phenotype space of the problem.

1.2.2 ^{13}C-MFA

^{13}C-MFA is a technique that combines experiments with computation for estimation of metabolic flux from ^{13}C-labeling experiments (Wiechert et al., 1999; Zamboni et al.,

2009). An organic substrate with strategically labelled ^{13}C atoms is fed to cells in ^{13}C-labeling experiments. The ^{13}C atoms are distributed through the metabolic network by the formation and breakage of bonds. Different metabolic pathways with different pattern of bond breakage results in different isotopic signatures (**Figure 1.2**) in metabolites (Wiechert et al., 1999; Zamboni et al., 2009). The isotope labeling patterns can be analyzed experimentally by techniques like nuclear magnetic resonance (NMR) or mass spectroscopy (MS) (Wiechert et al., 1999; Fischer et al., 2004). The use of MS involves the harvesting of cells after a certain time to allow the ^{13}C-label to be distributed (Zamboni et al., 2009). The constituents of biomass, e.g. proteinogenic amino acids, lipids etc., are extracted and purified. Additionally, the soluble metabolites may also be extracted and purified to determine their mass isotopomer distributions (MID) (Wiechert et al., 1999). A mass spectroscope determines the mass isotopomer distribution (MID) of different isotopomers of a metabolite or its fragments (Dieuaide-Noubhani and Alonso, 2014). A system of mass balance equations written in terms of isotopomer balances is solved by global optimization to estimate the fluxes (Wiechert et al., 1999). The fluxes can be estimated for both steady-state and nonstationary conditions (Kruger et al., 2011).

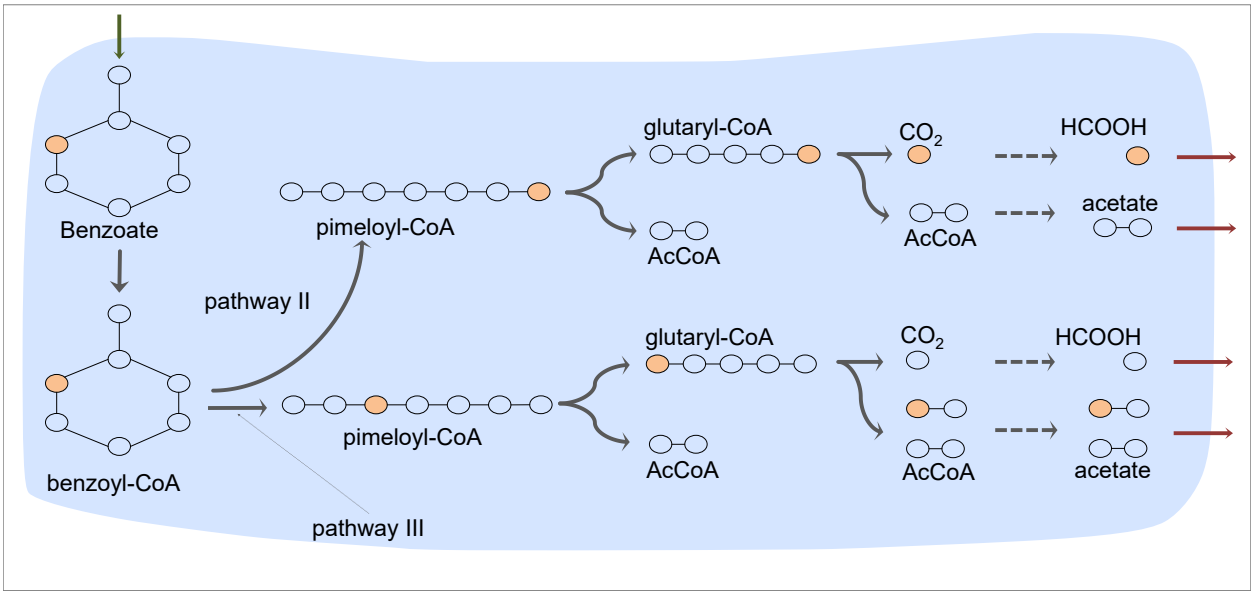


Figure 1.2. Schematics of ^{13}C-MFA relies on unique isotopic signatures of different pathways. Benzoate is degraded by two parallel metabolic pathways resulting in the same end products, but with differently positioned ^{13}C atoms in the end products and intermediates. The colored circles indicate ^{13}C-labelled atoms.

Compared to flux predictions of FBA, the fluxes estimated by ^{13}C-MFA are highly reliable. However, ^{13}C-MFA is experimentally and computationally intensive (Wiechert et al., 1999; Zamboni et al., 2009). The analysis of isotopomers requires expensive experimental setup like GC-MS or NMR with trained personnel. The mathematical treatment of ^{13}C-labeling experiments is significantly more complicated than the simple LP formulation of FBA. A metabolite with n number of carbon atoms has a total of 2^n isotopomers (Wiechert et al., 1999; Ganesh Sriram et al., 2004; Zamboni et al., 2009). An increase in the number of reactions, thus in metabolites results in an explosion in the number of parameters. Additionally, the balance equations are nonlinear in terms of the

fluxes and isotopomers. The solution of such a system of equation is computationally expensive, limiting the size of the network used. Multiple methods have been developed in the last 15 years to simplify the system of equations, e.g. cumomer (Wiechert et al., 1999), bondomer (Sriram, 2004), elementary metabolite units (Antoniewicz et al., 2007), to obtain a flux distribution with reduced computation time.

1.3 Metabolic Flux Analysis in Strain Development

Recently, metabolic flux analysis has seen significant employment in microbial strain development. The design process for strain improvement (**Figure 1.3**) has centered on *a priori* analysis of metabolic network of organisms as well as experimental verification of intercellular flux distribution, often in an iterative fashion. The common theme in flux guided strain design is the application of a predictive modeling technique like FBA to predict the genetic engineering target sites while ^{13}C-MFA is employed to analyze the performance of engineered strains. The rise in the popularity of flux guided strain design has been driven by the preceding development of software packages for automated reconstruction of genome-scale models and computational frameworks for analysis of the networks. The development of a completely automated workflow for construction of FBA ready plant genome-scale models without needing any curation have not been realized yet. However, predictive modeling and ^{13}C-MFA have seen significant usage in plant metabolic modeling community.

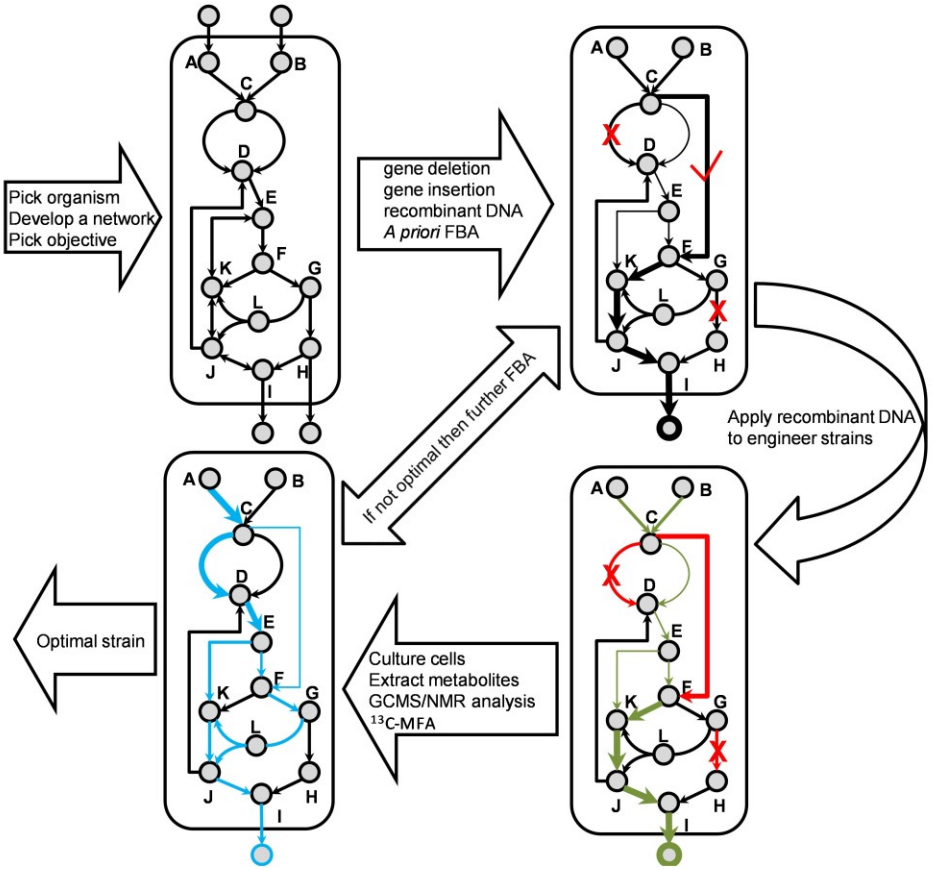


Figure 1.3. Overview of computational strain design. Flux balance analysis is employed to find targets for gene deletion or upregulation. Genes are upregulated or

deleted as per FBA predictions. The metabolic fluxes in the organism are estimated from ^{13}C-labeling experiments following the application of recombinant DNA technology. If the designed strain is suboptimal then the process is repeated. Circles indicate metabolites and arrows indicate reactions. Checkmarks imply gene upregulation while cross marks imply gene deletion, as predicted by FBA. Green arrows indicate increased flux and red arrows indicate decreased flux after genetic manipulation. Blue arrows indicate optimal flux equal to FBA prediction. Arrow thickness is proportional to metabolic fluxes the reactions carry.

A priori flux analysis of metabolic networks has generally focused on development of a phenotype with the highest yield. Commonly adopted strategies are cofactor balancing, gene deletion/addition, metabolite balancing, applied by themselves as standalone methods or in conjugation with one another. The computational complexity of *in silico a priori* flux analysis grows significantly with an increase in the number of genes considered for addition or deletion. Synthetic biology tools have made it a routine practice to add a non-native pathway or delete a redundant one. However, getting a non-native pathway to express sufficiently in a heterologous host itself is a major challenge. As a result, iterative approaches of trial and error are still prevalent in metabolic engineering (**Figure 1.3**).

1.4 Computational Methods in Predictive Modeling

Optimization is at the heart of predictive modeling of metabolism, with a rich history of successful applications. Over the last 25 years, the role of predictive modeling has progressed from simple *a priori* yield calculations with small networks to predicting metabolism with multi-tissue models spanning tens of thousands of reactions. The groundwork for predictive modeling methods was laid down in the form of development of flux balance analysis, elementary flux mode analysis and extreme pathway analysis to explore network topology and for usage in metabolic engineering (Bordbar et al., 2014). With the explosion of whole genome sequencing of organisms, application of these methods was extended to formulation of genome-scale metabolic models consisting of entire known inventory of reactions in an organism's genome. With advances in omics technologies, transcriptomics and metabolomics data started finding applications in predictive modeling frameworks to constrain the feasible solution space. Transcriptomics data have been widely used for guiding mode predictions as well as validating them (Bordbar et al., 2014). This has also been used in development of context specific metabolic models for plant and human tissues encompassing a subset of total inventory of metabolic reactions of the organisms.

A major development that was predictive modeling was addressing the limitation of single objective of linear programing based predictive modeling methods by development of bilevel optimization frameworks, in which a linear programing problem is employed as a constraint to an outer problem. ObjFind, the first published bilevel optimization framework, was developed for inferring hypothesized metabolic objective functions of organisms. Soon after the Optknock framework was developed for identifying optimal strain engineering strategies with the objective of finding out the minimum number of genetic interventions with minimal metabolic adjustment. Recently bilevel optimization

has been employed to integrate transcriptional regulatory networks and metabolic networks, investigate community dynamics of microbes, integrate gene expression data with genome-scale models etc. Much of this development has been driven with the specific goals of addressing some of the uncertainties associated with the limitations of strictly constraint-based analysis, and the objectives of the study. Additionally, methods have been proposed for designing novel metabolic pathways while accounting for thermodynamics and enzyme availability (Wang et al., 2017). Recently, macromolecular expression accounting for proteome synthesis (Lloyd et al., 2018) and whole-cell computational modeling (Karr et al., 2012) have also been demonstrated.

With incorporation of omics data and additional layers of complexity the predictive power improves significantly. However, this comes at the expense of increased computational demand. Linear programing problems can be solved in polynomial time. The existence of integer variables in many of the omics integrated methods make the problems NP hard. Additionally, very few solvers exist for solving linear bilevel optimization problems and cannot readily be used for large scale problems. Hence, the bilevel problems must be reduced to single level problems before they can be solved. Often, development of new algorithms is a necessity before solving certain classes of such problems, especially in the presence of integer variables. This is an active area of research in the field of operations research with some major developments in the last five years.

1.5 Challenges in Predicting Plant Metabolism

Challenges exist in prediction of plant metabolism due a multitude of factors. Plant metabolic networks contain significantly larger number of reactions and metabolites, compared to those of unicellular organisms, due to the presence of a huge array of metabolic pathways for secondary metabolite biosynthesis. Reactions, including entire pathways, are often duplicated in multiple organelles. Across different organelles, the metabolites are often trafficked by inter-organelle transporters, many of which have still not been identified. All these factors enhance the degree of interconnectivity within metabolic networks significantly, imparting a very large degree of freedom to plant metabolic networks. As a result, there are often an infinite number of equally optimal solutions to optimization problems with plant metabolic networks. This necessitates that the required level of scrutiny before accepting a flux solution as biologically relevant is significantly higher.

A challenging aspect of modeling plant metabolism is the lack of experimentally determined constraints that can be employed to cut down the solution space. Incorporation of data on nutrient uptake from cell culture media, product excretion to media and construction of biomass equation to the extent possible is ingrained in predictive modeling of microbes because of the relative ease of growing a culture. Plants have a wide variety of tissues designated to perform specific tasks, with vastly different metabolism. Hence, modeling a plant tissue requires tissue specific constraints on metabolite uptake and secretion to accurately mimics the metabolism. However, it is extremely challenging to obtain data on metabolite uptake and secretion for plant tissues. While suspension cell cultures can be obtained from the tissues of interest, they are often

not representative of actual metabolism in the tissues. As a result, plant metabolic modeling is often performed with a much smaller number of experimental constraints despite the larger size of the network.

The choice of objective function in plant metabolic modeling is not so obvious. Maximization of biomass synthesis in experimentally determined proportions is the generally the most commonly used objective function in plant metabolic modeling, often with minimization of the total reaction fluxes in the network. In addition to biomass synthesis, transport and cell maintenance costs are significant and requires incorporation in the objective function for accurately capture the metabolic demand of a growing plant cell. Construction of the biomass equation of plant cells is very challenging, especially in case of tissues that cannot be easily separated from a plant without destroying it, e.g. embryo and pollen. Within a tissue as well there exists spatiotemporal dynamics of metabolism, which impacts cellular objective and product synthesis, and needs to be addresses while performing predictive modeling.

All the aforementioned factors render predictive modeling of plant metabolism an inherently iterative process. In this process predictive modeling is performed with the limited constraints on the metabolic network and the results are analyzed carefully. After the biologically insensible predictions are identified, they are eliminated, and the predictive modeling process is repeated. This process is generally carried out until satisfactory experiment supported results are obtained. The flux solution is then inspected to obtain nontrivial information on metabolism. This is a slow tedious process, and better methodologies need to be developed to speed up this workflow.

1.6 Outline of Dissertation

The overreaching objective of this work is to employ mathematical optimization with metabolic networks to dissect hitherto unexplained biological phenomena and to provide metabolic justifications of the phenomena based on the optimization-based analyses. Optimization finds the best solution to an objective function under a given set of constraints, if the optimization problem is feasible. Application of optimization methods with biological networks has a rich history, primarily because of the ease of repeated analyses with a change in constraints or in objective functions, once the mathematical model representing the biological problem is constructed. Accurate measurements of nutrient uptake and product excretion act as constraints. The rate of biomass accumulation widely used as a proxy for growth rate of an organism. To date, investigations of cellular metabolism with optimization frameworks have been performed with significant amount of experimental data on nutrient uptake and product synthesis to constrain the solution space. We demonstrate that predictive modeling can be employed even in the absence of detailed nutrient uptake and product synthesis data, to yield biologically meaningful information by employing an approach based on (i) sensitivity analysis, (ii) analysis based on multiple metrics and (iii) development of novel computational frameworks. We have demonstrated the feasibility of this approach in our investigation of metabolism of a cellulosic bacterium and two plant systems.
The second chapter of this dissertation investigates the carbohydrate metabolism of the heterotrophic cellulosic bacterium *Saccharophagus degradans*. Sensitivity analysis of the

linear optimization problem capturing cellular growth of the organism on multiple organic substrates in flask cultures distinctly points towards amino acid fermentation as a coping mechanism under nutrient rich hypoxic coastal environment. The sensitivity analysis in the form of inspection of the dual or the primal linear optimization problem alleviates the limitation of a single level optimization formulation but retains the advantage of fast computational analysis.

The third chapter contains an approach based on construction of interacting tissue specific metabolic networks to parse information on cellular metabolism on the interacting tissues. We reconstructed a genome-scale metabolic network of *Populus trichocarpa* and employed flux balance analysis to analyze two tree specific metabolic events: (i) season internal nitrogen recycling by poplar trees, and (ii) emission of isoprene from poplar leaves on hot sunny days. The analyses based on multiple metrics provide a metabolic justification of internal nitrogen recycling, a phenomenon observed in deciduous trees. The predictions agree with experimental data reported in the literature, which further validated our computational predictions.

The fourth chapter we investigated the energetic need of maize endosperm metabolism during grain filling by reconstructing a genome-scale metabolic model of maize. We examine the following questions (i) What are the of contributions of all ATP generating reactions in fulfilling the energetic need of grain filling under hypoxia; (2) How endosperm metabolism responds under genetic perturbations; and (3) How flexible is the metabolic pathway during grain filling. The solution strategy for investigating the questions is based on flux balance and flux variability analyses, which are methods based on linear programming. The analyses predicted that under hypoxia, the endosperm is reliant on amino acid fermentation to metabolize carbohydrates and to generate ATP. The predictions are supported by experimental measurements. We hypothesize that the synthesis of storage proteins in the maize kernels is not only a way of storing nitrogen for seed germination, but also a way of optimizing the starch storage process.

The fifth chapter investigates the hypothesis developed in the previous chapter. We develop and employ bilevel optimization framework to investigate the temporal dynamics of storage protein accumulation in maize endosperm with a minimal set of data. We investigate the following questions: (i) Is the accumulation of different storage proteins result of limitations imposed by the structure of the metabolic network; and (ii) What is the role of nutrient profile in dictating the accumulation of storage proteins. The analyses predict the onset of accumulation of different storage proteins accurately.

Chapter 2 : Investigation of Carbohydrate Metabolism of a Heterotrophic Cellulosic Bacterium

This chapter is part of a manuscript under preparation. Copyright maybe transferred to the publisher in the future.

Navadeep Boruah, Brendon K. Cox, Steven H. Hutcheson and Ganesh Sriram*

NB, BKC and GS conceived this project. NB and BKC constructed and analyzed the GSM, GS provided input on analyses, SHH provided input on *Sde* metabolism. NB and BKC wrote the manuscript, GS and SHH critically edited it. N.B. and B.K.C. contributed equally to this work.

Department of Chemical and Biomolecular Engineering (N. B., B. K. C., G. S.) and Department of Cell Biology and Molecular Genetics (S. H. H), University of Maryland, College Park, MD 20742, United States of America

2.1 Background

The ocean is the largest CO_2 sink (Riebesell et al., 2007). Autotrophic organisms fix CO_2 into organic carbon, and heterotrophic organisms convert this organic carbon back to inorganic carbon (Hügler and Sievert, 2011). This balance between heterotrophy and autotrophy plays a key role in global carbon cycle (Smith and Hollibaugh, 1993), and regulates atmospheric CO_2 and O_2 concentrations (Hügler and Sievert, 2011), thus playing a role in climate change (Bauer et al., 2013). The global marine carbon cycle is not yet fully understood (Bauer et al., 2013). The presence of many major contributing pools (Bauer et al., 2013), fluxes with spatiotemporal dynamics (Bauer et al., 2013), uncertainties associated with the carbon fluxes (Churkina, 2016) and still emerging roles of the coastal subsystems in the carbon cycle (Bauer et al., 2013) render it highly complex. Uncertainties associated with carbon fluxes from coastal ecosystems (Bauer et al., 2013) have made the role of coastal ecosystems unclear (Kathilankal et al., 2008; Kirwan and Mudd, 2012). A thorough understanding of fluxes, and factors regulating them, are essential to accurate prediction of carbon cycle dynamics (Bauer et al., 2013).

Tidal and salt marshes, major subsystems of the coastal ecosystem, contain significant reserves of organic carbon as lignocellulosic biomass (Duarte et al., 2005; Gontikaki et al., 2015; Leithold et al., 2016). Conversion of organic carbon to inorganic form in these ecosystems is driven primarily by microbes (Churkina, 2016; Leithold et al., 2016), prominently *gammaproteobacteria* (Dyksma et al., 2016). Additionally, utilization of lignocellulosic biomass as substrates for large scale chemical production is a major goal of metabolic engineering. Thus, thorough understanding of *gammaproteobacteria* metabolism can improve accurate modeling of carbon cycle in coastal subsystems and advance the biotechnology industry.

2.1.1. *Saccharophagus degradans* a model aerobic cellulosic microbe

So far, multiple organisms capable of metabolizing cellulose, a beta-1,4-glucan polymer, have been identified and their genome sequenced. However, very few organisms can degrade lignin, which poses a major challenge to the biological processing of lignocellulosic biomass. *Saccharophagus degradans* strain 2-40 (Sde 2-40), isolated from decaying salt marsh grass in Chesapeake Bay (Andrykovitch and Marx, 1988; Weiner et al., 2008), is an aerobic cellulosic bacterium with putative ligninases (Weiner et al., 2008). Sde 2-40, henceforth referred to as *Sde*, was the first aerobic cellulosic bacterium to have its genome sequenced (Weiner et al., 2008). We employed *Sde* as a model organism to assess carbohydrate metabolism in the context of constraints imposed by nutrient availability in coastal environment.

2.2 Materials and Methods

2.2.1 Construction of Draft Model

We reconstructed the first draft of the genome-scale metabolic model (GSM) by mining the data files from the *BioCyc* database. The raw data files were mined by using a host of internally developed MATLAB scripts. This draft model, with 1449 reactions and 1022 metabolites, was not FBA-ready. Hence, the GSM was curated manually in an iterative fashion (**Figure 2.1**) to: (i) fill the gaps in the network; (ii) remove substrate cycles and (iii) remove all generic reactions and dead-end metabolites. Finally, a biomass synthesis reaction, originally constructed for *E. coli* (Reed et al., 2003), was assigned to the cytosol.

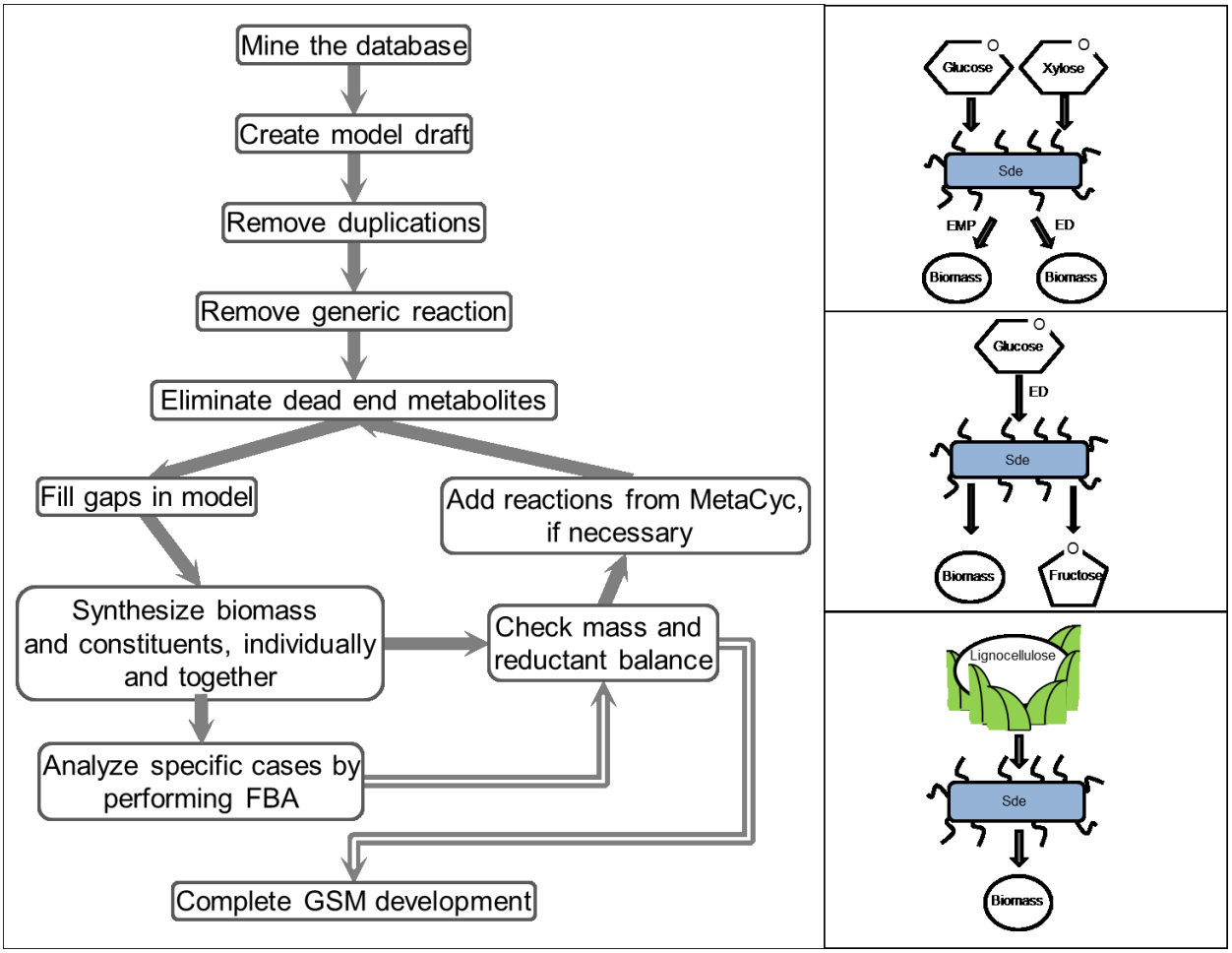


Figure 2.1. Workflow of genome-scale model construction of *S. degradans* strain 2-40. The first draft of the model was constructed by mining the *BioCyc* data file of the organism with MATLAB scripts developed in-house. The model draft was then curated manually, as well as MATLAB scripts, to remove duplications of reactions and metabolites, generic reactions and dead-end metabolites. A cyclical model curation method was then employed to find gaps in the model and to ensure removal of substrate cycles.

2.2.2 Gap Filling

We identified metabolic gaps in the GSM by performing the FBA for biomass synthesis. We fixed the gaps based on the following evidence, in the following order of preference: (i) addition of reactions with experimental evidence of existence in *gammaproteobacteria*; (ii) addition of reactions from any known bacterial pathway in *MetaCyc*, if all the reactants and products of a reaction were already present in the model; (iii) addition of putative reactions in bacteria for the sole purpose of assessment of metabolic potential with genetic modification of *Sde*. We assumed that *Sde* has the transporters necessary to take up all the nutrients essential to the FBA analyses performed in this work and incorporated them to our model as extracellular transport reactions.

2.2.3 Substrate Cycles

The substrate cycles in the model led to mass and energy imbalances leading to unrealistic production of biomass and its precursors. Majority of substrate cycles contributing to mass imbalances got eliminated upon correcting elemental mass balance of reactions. The substrate cycles producing energy imbalances, by generating excess ATP and reductant, were removed by constraining directionalities of the reactions based on Gibbs' free energy. To the best of our knowledge, the GSM is completely free of substrate cycles.

2.2.4 Generic Reactions and Dead-End Metabolites

Generic reactions with R-groups in the metabolites were replaced with reactions involving specific metabolites. Often a representative metabolite was used instead of multiple specific metabolites. Then the reactions were removed in the absence of any specific metabolite. The pathways involving polymerization of a monomer, for example starch biosynthesis, were replaced with a single reaction with the monomer as the sole reactant and the polymer as the sole product. The dead-end metabolites were identified by analyzing the stoichiometric matrix of the model and were removed from the model.

2.2.5 Mathematical Formulation

FBA uses linear programming to solve the following problem:

$$\begin{aligned} \max \quad & c_j v_j \\ \text{s.t.} \quad & S_{ij} v_j = 0 \\ & v_{j,min} \leq v_j \leq v_{j,max} \end{aligned}$$

This is the primal of the linear programming problem. The FBA simulations were performed by using the COBRA toolbox in a MATLAB environment with the open source GNU Linear Programming Kit (GLPK) as the solver used in all the simulations. The minimization of L_1 norm was the secondary objective in all analyses.

2.3 Results

The catabolism of carbohydrates in bacteria takes place by one or a combination of the three following pathways: (i) Emden-Meyerhof-Parnas (EMP) pathway, (ii) Entner-Doudoroff (ED) pathway and (iii) pentose phosphate pathway (PPP). The EMP pathway converts glucose (Glc) to pyruvate (Pyr). The ED pathway converts it to triose phosphates (GAP) and pyruvate (Pyr). GAP produced by ED pathway is further converted into pyruvate through the remainder of the EMP pathway (Chen et al., 2016). Presence of EMP pathway in eukaryotes is ubiquitous. However, prokaryotes generally display much more diversity (Flamholz et al., 2013). The ED pathway is generally present in bacteria inhabiting harsh nutrient conditions (Klingner et al., 2015). Earlier studies have reported that 90% marine bacteria employ the ED pathway as the preferred pathway for glucose utilization, despite generation of less ATP per mole of substrate metabolized (Klingner et al., 2015). The genome of *Sde* contains putative genes of both EMP and ED pathways. It is unclear if *Sde* employs both the pathways for carbohydrate utilization or if the presence of both pathways confers it any metabolic advantage.

2.3.1 Metabolism of Glucose, Cellobiose and Xylose by *Sde*

Carbohydrate metabolism of *Sde* involves excretion of endoglucanases and glucosidases to degrade cellulose outside of the cell to cellobiose (Zhang and Hutcheson, 2011). Cellobiose is then metabolized by *Sde* through beta-glucosidases on the surface and inside the organism (Zhang and Hutcheson, 2011). Glucose, cellobiose and xylose are major saccharides of cellulose. Experiments have confirmed that *Sde* can grow on glucose, cellobiose and xylose. *Sde* excreted an unconfirmed hexose carbohydrate when grown on cellobiose and glucose in flask cultures (unpublished data). Hexose excretion appears wasteful, and the reason is unclear. ^{13}C-MFA experiments in our lab have shown that *Sde* employs ED pathway for carbohydrate metabolism (unpublished data). This is counterintuitive, as ED pathway should not be utilized at all if the organism can utilize the more efficient EMP pathway.

We analyzed *Sde's* growth on these three substrates with FBA and observed that *Sde* preferentially employed EMP pathway with all substrates. We investigated if *Sde* employs ED pathway due to any metabolic bottleneck in EMP pathway. We placed flux restrictions on enzymes of EMP pathway and repeated the analyses. By restriction of EMP pathway we specifically mean restriction of a reaction step present in only EMP pathway. Additionally, the term restriction implies a decreased non-zero bound on reaction flux. Blocking implies complete elimination of a reaction with flux bounds of zero. Predicted biomass yield decreases upon restriction of EMP pathway reactions on all three substrates.

We blocked EMP pathway by blocking reactions catalyzed by the enzymes *phosphofructokinase (pfk)*, *glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*) and *phosphoglycerate kinase (pgk).* ED pathway carried flux when *pfk*, the enzyme phosphorylating fructose-6-phosphate (F6P), was blocked (**Figure 2.2, Figure 2.3**). We investigated the differences in the central carbon metabolism of *Sde*, on Glc and Xyl, with *pfk* restriction. *Pfk* blockage was associated with decreased folate metabolism and higher TCA cycle fluxes (**Figure 2.2, Figure 2.3)**. With a blocked EMP pathway, *Sde* on xylose produced more biomass per unit carbon than on glucose. The PPP flux was minimal on Glc (**Figure 2.2**), and substantial on Xyl (**Figure 2.3**). The results suggest an impact of limitations of enzymatic capacity of key reactions on the preferred nutrient of *Sde*.

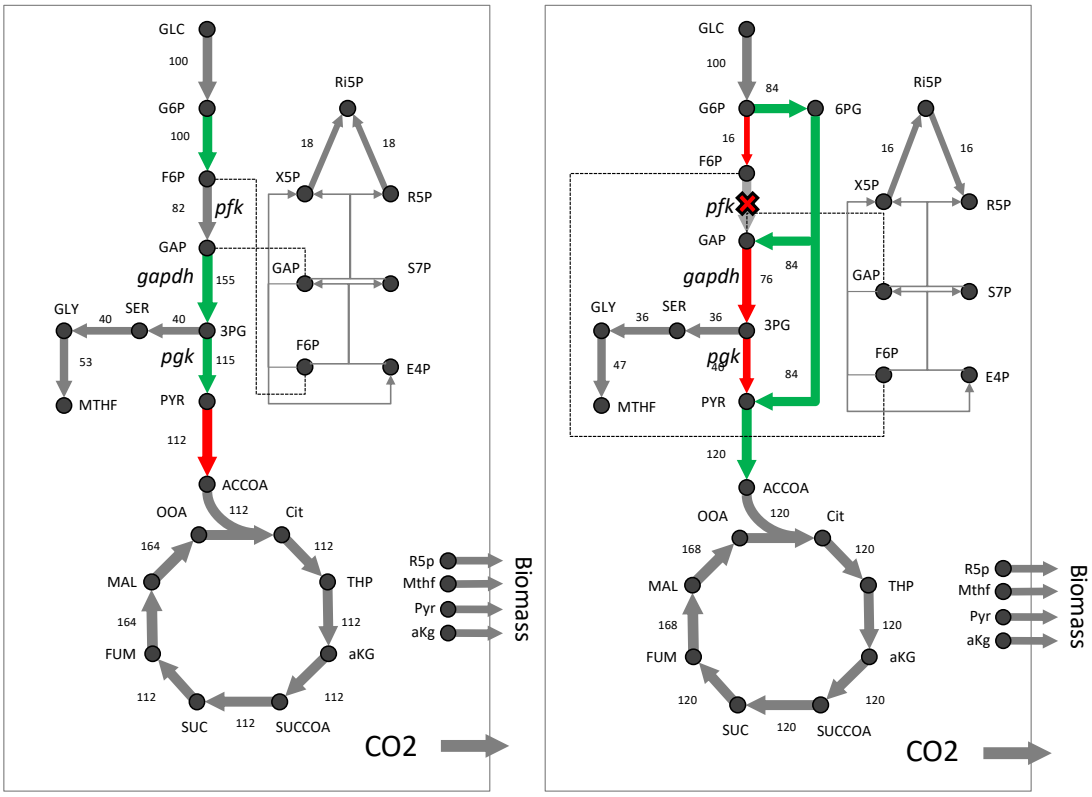


Figure 2.2. Altered flux distributions of *Sde* on Glc predicted upon blockage of *pfk*. We analyzed the impact of blocking *pfk* on central carbon metabolism of *Sde* with Glc as the substrate. FBA employed the objective function of biomass maximization with minimization of L1 norm, with 100 moles of Glc as substrate and unrestricted inorganic nutrients uptake. Circles represent metabolites and solid arrows represent reactions with the directions. For the sake of brevity, only key reactions and metabolites are shown. Arrow thickness represents reaction flux in a logarithmic scale, with the flux next to the arrow. Dotted arrows connect identical metabolites, for ease of display, shown in different places on the flux map. The red "X" denotes a blocked reaction. All abbreviations of metabolites are defined in the List of Abbreviations. A, flux map of *Sde* with active *phosphofructokinase* enzyme. B, flux map of *Sde* with blocked

phosphofructokinase enzyme.

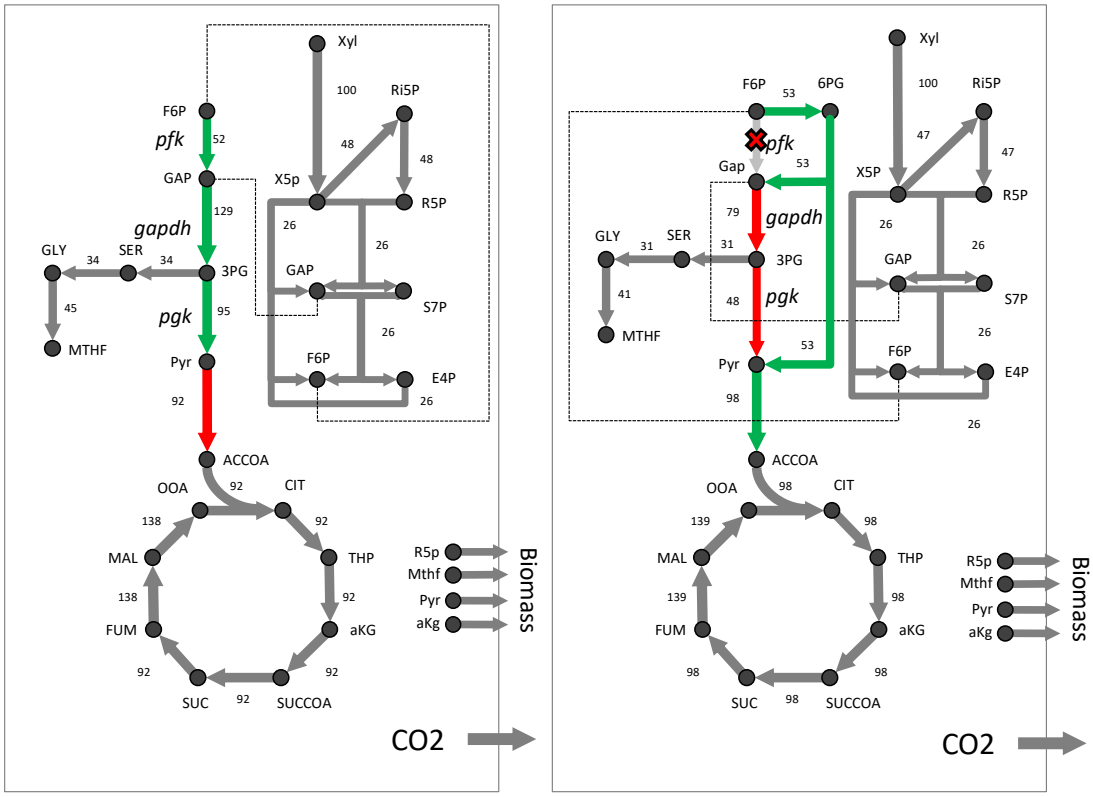


Figure 2.3. Altered flux distributions of *Sde* on Xyl predicted upon blockage of *pfk*. We analyzed the impact of blocking *pfk* on central carbon metabolism of *Sde* with Xyl as the substrate. FBA employed the objective function of biomass maximization with minimization of L1 norm, with 100 moles of Xyl as substrate and unrestricted inorganic nutrients uptake. Circles represent metabolites and solid arrows represent reactions with the directions. For the sake of brevity, only key reactions and metabolites are shown. Arrow thickness represents reaction flux in a logarithmic scale, with the flux next to the arrow. Dotted arrows connect identical metabolites, for ease of display. The red "X" denotes a blocked reaction. All abbreviations of metabolites are defined in the List of Abbreviations. A, flux map of *Sde* with active *phosphofructokinase* enzyme. B, flux map of *Sde* with blocked *phosphofructokinase* enzyme.

When ED pathway was used for carbohydrate utilization, uptake of O_2 increased significantly (~50%) (**Figure 2.4**). A cycle involving glutamate dehydrogenase, which uses 2-ketoglutarate and ammonia to produce glutamate, and aspartate aminotransferase, which converts glutamate and oxaloacetate into aspartate and 2-ketoglutarate (**Figure 2.4**) was predicted to be active. Asp was converted into fumarate through adenylsuccinate lyase (**Figure 2.4**).

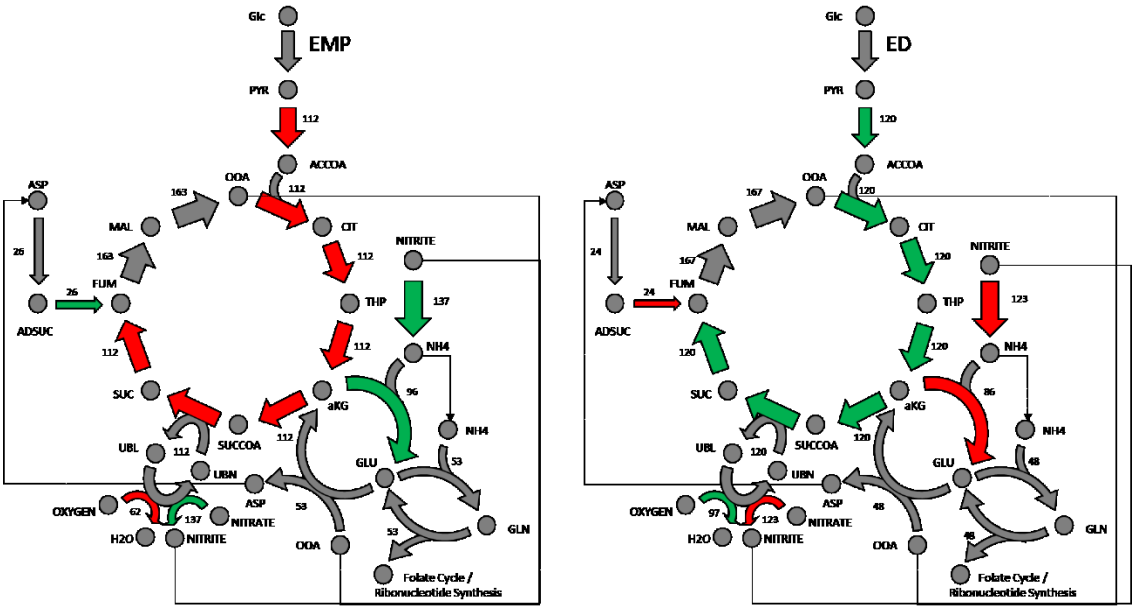


Figure 2.4. TCA cycle fluxes in *Sde* vary depending upon EMP vs ED pathway activity. Key differences in fluxes of TCA cycle were observed depending upon whether ED or EMP pathways was in operation. With EMP pathway in operation, TCA cycle reactions carried smaller fluxes than with ED pathway in operation. Higher flux in TCA cycle with ED pathway in operation results in a higher oxygen demand. Additionally, reactions involved in nitrogen assimilation carry higher fluxes when EMP pathway is used in assimilation of glucose. Circles represent metabolites and solid arrows represent reactions with the directions. For the sake of brevity, only key reactions and metabolites are shown. Arrow thickness represents reaction flux in a logarithmic scale, with the flux next to the arrow. Dotted arrows connect identical metabolites, for ease of display, shown in different places on the flux map. The red "X" denotes a blocked reaction. All abbreviations of metabolites are defined in the List of Abbreviations. A, metabolic flux distribution in TCA cycle and its associated reactions in *Sde* with EMP pathway as the preferred pathway for glucose utilization. B, metabolic flux distribution in TCA cycle and its associated reactions in *Sde* with ED pathway as the preferred pathway for glucose utilization.

Sde was originally isolated from an environment rich in carbon nutrients. However, in coastal ecosystems organisms generally experience inorganic nutrient limitations (Van Weijnen et al., 1999). We suspected that hexose excretion by *Sde* could be due to limitation of enzymatic capacity to process the organic substrates inside the cell. To investigate the effects of enzymatic limitation, we performed FBA with blocked *pfk* to initiate the ED pathway, with Glc as the substrate. Additionally, we restricted *2-dehydro-3-deoxy-phosphogluconate aldolase* (*kdpg*), the last reaction step in the ED pathway, and *6-phosphogluconate dehydrogenase* (*6pgd*), the last step in oxidative PPP. The resulting simulation predicted that the organism would excrete fructose (**Figure 2.5**). When fructose excretion was not allowed, FBA could not find a solution. We repeated the analyses with blocked *gapdh* or *pgk* instead of *pfk*. In these simulations the model did not excrete fructose. Instead, the model used EMP pathway synthesized pyruvate through

dihydroxyacetone phosphate (**Figure 2.5**). FBA predicted similar results with cellobiose as the substrate. The model, with xylose as the substrate, could not find a solution when *pfk* was blocked.

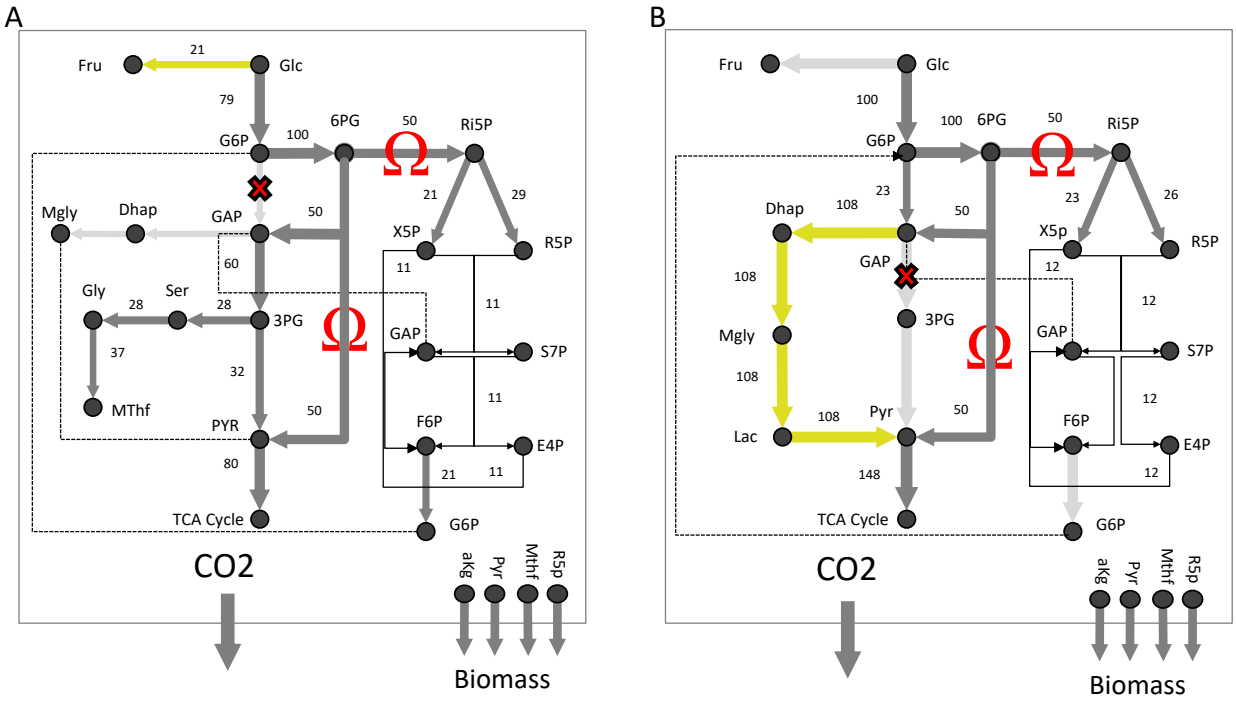


Figure 2.5. Flux distributions of *Sde* blocked EMP pathway enzymes (*pfk*, *gapdh* or *pgk*) and partial restriction of ED pathway and PPP. Experiments have shown that *Sde* excretes an unconfirmed hexose carbohydrate when it grows on glucose and cellobiose. FBA predicted that *pfk* deficient *Sde* on Glc as substrate will excrete Fru if on ED pathway and PPP fluxes are restricted. All the simulations employed the objective function of biomass maximization, with minimization of L1 norm, on 100 moles of Glc and unrestricted inorganic nutrients uptake. Circles represent metabolites and solid arrows represent reactions with the directions. For the sake of brevity, only key reactions and metabolites are shown. Arrow thickness represents reaction flux in a logarithmic scale, with the flux next to the arrow. Dotted arrows connect identical metabolites, for ease of display. The red "X" denotes a blocked reaction. The red ohm symbols represent partial flux restrictions placed on the model. An arrow that is shaded yellow represents a new flux path that the model takes when the ED and PPP reaction restrictions are placed. All abbreviations of metabolites are defined in the List of Abbreviations. A, predicted flux profile in *Sde* with blocked *phosphofructokinase* (*pfk*), restricted *KDPG aldolase* (*kdpg*) and restricted *6-phosphogluconate dehydrogenase* (*6pgdh*). B, predicted flux profile in *Sde* with blocked *glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*), restricted *KDPG aldolase* (*kdpg*) and restricted *6-phosphogluconate dehydrogenase* (*6pgdh*).

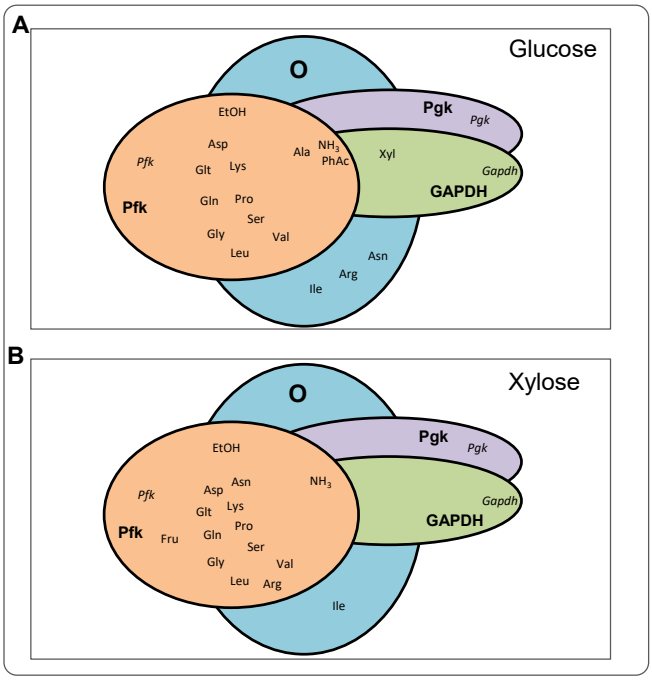


Figure 2.6. Shadow prices of internal and extracellular reactions under enzymatic limitation and hypoxia. We performed FBA to identify reactions that could limit *Sde's* growth under hypoxia. The simulations employed the objective function of biomass maximization with blocked O_2 uptake, unrestricted uptake of other nutrients, and blockage of one or more EMP pathway reactions. We analyzed the shadow prices of reactions to identify the limiting reactions. Excretion reactions with positive shadow prices indicate elevated biomass synthesis upon relaxation of the constraint. The metabolites included in an oval indicate that their excretion could enhance biomass synthesis of *Sde*. Each oval indicates a simulation with a blocked EMP pathway enzyme, with the enzyme name in bold letters inside the oval. The metabolites contained within intersections of ovals indicate the excretion reactions with multiple blocked enzymes. All abbreviations of metabolites are defined in the List of Abbreviations.

2.3.2 Phosphate Limitation Dictate Pathway Selection

EMP pathway converts one molecule of Glc to two molecules of GAP, which are then converted to two molecules of Pyr. ED pathway converts one molecule of Glc to one molecule of Pyr and GAP. The conversion of GAP to Pyr employs reactions that involve inorganic phosphate and phosphorylation. Thus, by employing ED pathway, *Sde* reduces GAP synthesis. It had been reported that bacterial communities in marshes are primarily limited by phosphates, followed by labile carbon (Sundareshwar et al., 2003). Similarly, under phosphate starvation, plants bypass steps that require ATP or phosphate (Vance et al., 2003). Thus, utilization of ED pathway may be beneficial to *Sde* from the point of reducing the demand for phosphates. Organisms using ED pathway as the primary pathway do not have *pfk* encoded in their genome (Chen et al., 2016). Since Fru is excreted by *Sde*, inorganic nutrient utilization takes precedence over efficient utilization of fixed carbon.

2.3.2 Dual Analysis Points to Possible Amino Acid Fermentation

With ED pathway in operation in *Sde* increased TCA cycle fluxes were predicted, resulting in elevated oxygen uptake (**Figure 2.4**). The Chesapeake Bay contains severely hypoxic low-oxygen zones. This indicates that *Sde* potentially is not impacted by hypoxia. We analyzed the dual of the primal FBA problems with restricted oxygen uptake. We observed that the shadow prices of multiple amino acid excretion reactions (**Figure 2.6**) were positive. The indicated that fermentation under such nutrient conditions would increase biomass synthesis of Sde. Fermentation by bacteria under hypoxia is well known. However amino acid fermentation in marine aerobic heterotrophs like *Sde* needs further investigation.

2.4 Conclusion

We reconstructed a genome-scale metabolic model of *Saccharophagus degradans* strain 2-40 and employed FBA to investigate carbohydrate metabolism of the organism. The simulations indicated that phosphate limitations could dictate selection of carbohydrate degrading pathway. Additionally, analysis of the dual of the FBA problem predicted amino acid fermentation under hypoxia to support continual degradation of carbohydrates. Essentially, these results suggest that the dynamic intertwined nature of carbon and nitrogen cycle in coastal environment could be impacted significantly by bacterial metabolism. To assess the role of *gammaproteobacteria* metabolism on these cycles, an experimental investigation of carbohydrate metabolism by *Sde* is warranted.

Chapter 3 : A Genome-Scale Stoichiometric Model of *Populus trichocarpa* for Investigation of Tree Metabolism: Seasonal Nitrogen Mobilization and Isoprene Emission

This chapter is part of a manuscript under preparation. Copyright maybe transferred to the publisher in the future.

Navadeep Boruah, Ashish Misra, Margaret S. Senftle, Gary D. Coleman and Ganesh Sriram[★]

AM, NB and GS conceived the study; GDC provided critical insights on poplar biology; MSS created the first draft of the model and identified the organellar location of the enzymes; NB curated the model, created its final draft and performed FBA; NB and GS wrote the manuscript.

[*]Principal investigator and corresponding author

Department of Chemical and Biomolecular Engineering (N. B., A. M., M. N. S., G. S.) and Department of Plant Science and Landscape Architecture (G. D. C.), University of Maryland, College Park, MD 20742, United States of America

Present addresses: [†]A. Misra: Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, New Delhi, India; [‡]M. S. Senftle: ExxonMobil, Houston, Texas, United States of America

3.1 Background

The metabolic biology of woody plants (trees) is distinct from that of herbaceous plants. Trees have adapted to survive perennially and for longer timescales, as evidenced by wood formation (Taylor, 2002). In addition to large-scale CO_2 fixation, their metabolic capabilities include seasonal nitrogen cycling (Millard and Grelet, 2010) and volatile organic compounds (VOC) emission (Kesselmeier and Staudt, 1999). Seasonal nitrogen cycling involves the systematic transport of nitrogen from the shoots to the bark during leaf senescence in fall and in the reverse direction during leaf regrowth in spring (Babst and Coleman, 2018). VOC emission involves synthesis and emission of hydrocarbons such as isoprene into the atmosphere. A fundamental investigation of nutrient flow and partitioning in the underlying metabolic networks would advance plant biology and enable the biotechnological development of engineered trees.

For cell suspensions and small herbaceous plants, such investigations have been made possible by genome-scale metabolic models (GSM) and flux balance analysis (FBA) (e.g. Baghalian et al., 2014; Cheung et al., 2014; Dal'Molin et al., 2015; Levering et al., 2016), kinetic modeling (e.g. Rios-Estepa and Lange, 2007; Colón et al., 2010) and isotope-assisted metabolic flux analysis (MFA) (Young et al., 2011; Szecowka et al., 2013; Jazmin et al., 2014; Zhang et al., 2018). Of these, kinetic modeling and isotope-assisted MFA are experimentally intensive and therefore currently difficult to apply to the complex, seasonal metabolism of whole trees. GSM reconstruction may serve as a tool to

develop hypotheses to guide experiment design to facilitate investigation of metabolism and its regulation.

In the past decade, researchers have reported the reconstruction of quite a few plant GSMs and their analysis toward a range of metabolic predictions (Dal'Molin and Nielsen, 2018). Poolman et al. first published a genome-scale model for a leafy plant, *Arabidopsis thaliana*. Subsequently, multiple research groups independently published Arabidopsis GSMs by incorporating increasing levels of compartmentalization, inter-organellar transport and tissue specificity (Dal'Molin et al., 2010a; Saha et al., 2011; Mintz-Oron et al., 2012; Cheung et al., 2014) (Dal'Molin et al., 2010a; Saha et al., 2011; Mintz-Oron et al., 2012; Cheung et al., 2014). For example, the first fully compartmentalized Arabidopsis GSM, AraGEM, was used to investigate the interplay between photosynthesis and photorespiration (Dal'Molin et al., 2010a). GSMs for the C_4 plants maize, sugarcane and sorghum (Dal'Molin et al., 2010b; Saha et al., 2011; Simons et al., 2014b; Simons et al., 2014a; Seaver et al., 2015; Bogart and Myers, 2016), and the C_3 plants rice (Poolman et al., 2013), rapeseed (Pilalis et al., 2011), tomato (Yuan et al., 2016b), foxtail millet (Dal'Molin et al., 2016) have also been developed and analyzed to study metabolic scenarios ranging from photosynthesis to nitrogen assimilation. Recently, investigations with plant GSMs have been extended to multi-tissue analyses (Grafahrend-Belau et al., 2013; Dal'Molin et al., 2015). The predictions in most of these studies generally agreed with experimental findings, substantiating the power of GSMs analysis as a tool to investigate plant metabolism.

Poplar (genus *Populus*) consists of approximately 30 species of fast-growing trees. *P. trichocarpa*, commonly known as black cottonwood or California poplar, was the first tree to have its genome completely sequenced and published (Taylor, 2002; Tuskan et al., 2006; Sannigrahi et al.). Its modest genome size (~500 Mbp) and the availability of experimental tools for its genetic modification (Tuskan et al., 2006) make it an excellent platform for the development of a GSM to study metabolic questions unique to tree metabolism, which could be used as a foundation to design experiments. Here, we present the reconstruction of a GSM for *P. trichocarpa* (**Figure 3.1.**), PoplarGSM1.0, which can be readily used with COnsrtaints Based Reconstruction and Analysis (COBRA) toolbox (Becker et al., 2007; Schellenberger et al., 2011). PoplarGSM1.0 is fully compartmentalized, with reactions distributed amongst cytosol, extracellular, plastid, mitochondria and peroxisome **(Figure 3.2)**. We used this GSM to investigate two metabolic scenarios specific to tree metabolism: (i) seasonal nitrogen cycling during leaf senescence in fall and regrowth in spring, and (ii) emission of isoprene by Poplar on hot days.

3.2 Construction and curation of PoplarGSM1.0

3.2.1 Generation of raw model draft

We drafted an initial GSM containing over 2200 reactions with around 1900 metabolites by processing information in the PoplarCyc database (a part of PlantCyc) (Zhang et al., 2010) via the automated toolbox SuBliMinaL (Swainston et al., 2011). Such draft models generated by automated toolboxes are not FBA-ready and need extensive curation. Previous publications have discussed methodologies for using automated toolboxes for

GSM generation and subsequent curation to render the models FBA-ready (Seaver et al., 2012; Simons et al., 2014a). We followed the procedures described by us (Simons et al., 2014a) as well as the strategies described in below. **Figure 3.1.** depicts our workflow for model reconstruction, refinement and FBA.

3.2.2 Assigning subcellular compartmentalization to reactions

The draft genome-scale model incorporates available experimental or computational evidence for the compartmentalization for many reactions. To confirm the accuracy of the localization of a reaction, we looked up the localization of the Arabidopsis gene ortholog whose product catalyzes the reaction in three databases: (i) the Arabidopsis peroxisome (AraPerox) database (Reumann et al., 2004), (ii) the Arabidopsis chloroplast (AT_Chloro) database (Ferro et al., 2010), and (iii) the Arabidopsis mitochondrial protein (AMPDB) database (Heazlewood and Millar, 2005). Additionally, we used the algorithms TargetP (Emanuelsson et al., 2000) and WoLF PSORT (Horton et al., 2007) to predict subcellular location of a protein based on presence of signal peptide in the protein sequence. When any of these methods was able to predict a subcellular compartment for a reaction, we assigned to the reaction the compartment predicted by majority of these resources. When none of these methods yielded a result, we turned to previously published literature to assign the compartment. Failure to assign a compartment by any of these methods resulted in the reaction being assigned to cytosol. Spontaneous reactions that did not require an enzyme were assigned to all compartments. We repeated this process for all reactions in the model to arrive at a fully compartmented model. Finally, the biomass synthesis reaction was assigned to the cytosol and the extracellular transport reactions were added to the model. The inter-organelle transport reactions, which were added after finishing the compartmentalization, are the discussed in the following paragraphs. **Figure 3.2** shows the final distribution of metabolites and reactions in PoplarGSM1.0.

3.2.3 Gap filling

The fully compartmented GSM had multiple gaps and was not FBA-ready. We identified and filled the gaps in the model by performing the FBA simulations for biomass production in poplar cell suspensions as well as for our case studies. Failed simulations pointed to gaps in the metabolic network. A close inspection of the model revealed that the gaps were either due to missing metabolic reactions or missing inter-organelle transport reactions. For example, amino acid degradation pathways differ substantially from their biosynthesis pathways, and PoplarCyc does not contain the complete degradation pathways of several amino acids. We filled such metabolic gaps were filled by incorporating, in order of preference: (i) reactions mentioned in previously published literature on poplar, (ii) putative reactions from the Gramene database (Ware et al., 2002) and (iii) putative reactions from MetaCyc (Karp et al., 2002).

Inter-organellar transport reactions constituted major gaps in our compartmentalized model. Because information on the relevant transporter proteins in plants is often scarce, an optimization-based strategy became necessary to fill the corresponding gaps. Previously published compartmentalized plant genome-scale GSMs have used approaches based around optimization-based gap filling algorithms such as GapFill or

GapFind to fill these gaps (Kumar et al., 2007; Saha et al., 2011). This automated addition of inter-organellar transport reactions predicted over 800 transporters; however, we did not attempt to bioinformatically identify the genes encoding these transporters. A caveat of adding too many inter-organellar transport reactions is that it tremendously increases the number of degrees of freedom of the metabolic network, thus potentially leading to unrealistically large solution spaces. To avoid this situation, we kept the number of transport reactions to a minimum by removing all transport reactions that metabolites appearing as reactants (and products in reversible reactions) in exactly one organelle. Additionally, we identified the nonessential several transport reactions that carried zero flux in all our FBA simulations, and eliminated these from our model.

3.2.4 Generic reactions and dead-end metabolites

The draft GSM had 400 dead-end metabolites. They were fixed either (i) by finding reactions in literature which were not present in PoplarCyc or (ii) by removing the metabolites from the metabolic network. The generic reactions involving R-groups in the metabolites were replaced with reactions involving specific metabolites. If no specific reactions were found, the reactions were removed from the model. All pathways involving polymerization of a monomer in repeated steps, for example starch biosynthesis, were replaced with a single reaction which contains the monomer as the sole reactant and the polymer as the sole product.

3.2.5 Substrate cycles

The steps mentioned above rendered the model FBA-ready. However, the model generated inaccurate results due to the presence of substrate cycles, which were identified by mass balance calculations. Most of the substrate cycles were removed by ensuring elemental mass balancing of all the reactions in our model. However, several substrate cycles producing an energy imbalance were observed while analyzing the results of the FBA simulations. The substrate cycles generating ATP and reductant were removed by making some of the reactions involved in the reaction irreversible. To the best of our knowledge, the poplar GSM presented here is completely free of substrate cycles.

3.2.6 Development of the biomass reaction

A biomass equation was developed based on experimentally observed biomass composition of the suspension cells. The constituents of the biomass equation are proteinogenic amino acids, nucleotides, lipids, carbohydrates, lignin and cellulose. A biomass equation constructed in a similar manner has been used in another study from our groups (Nargund et al., 2014).

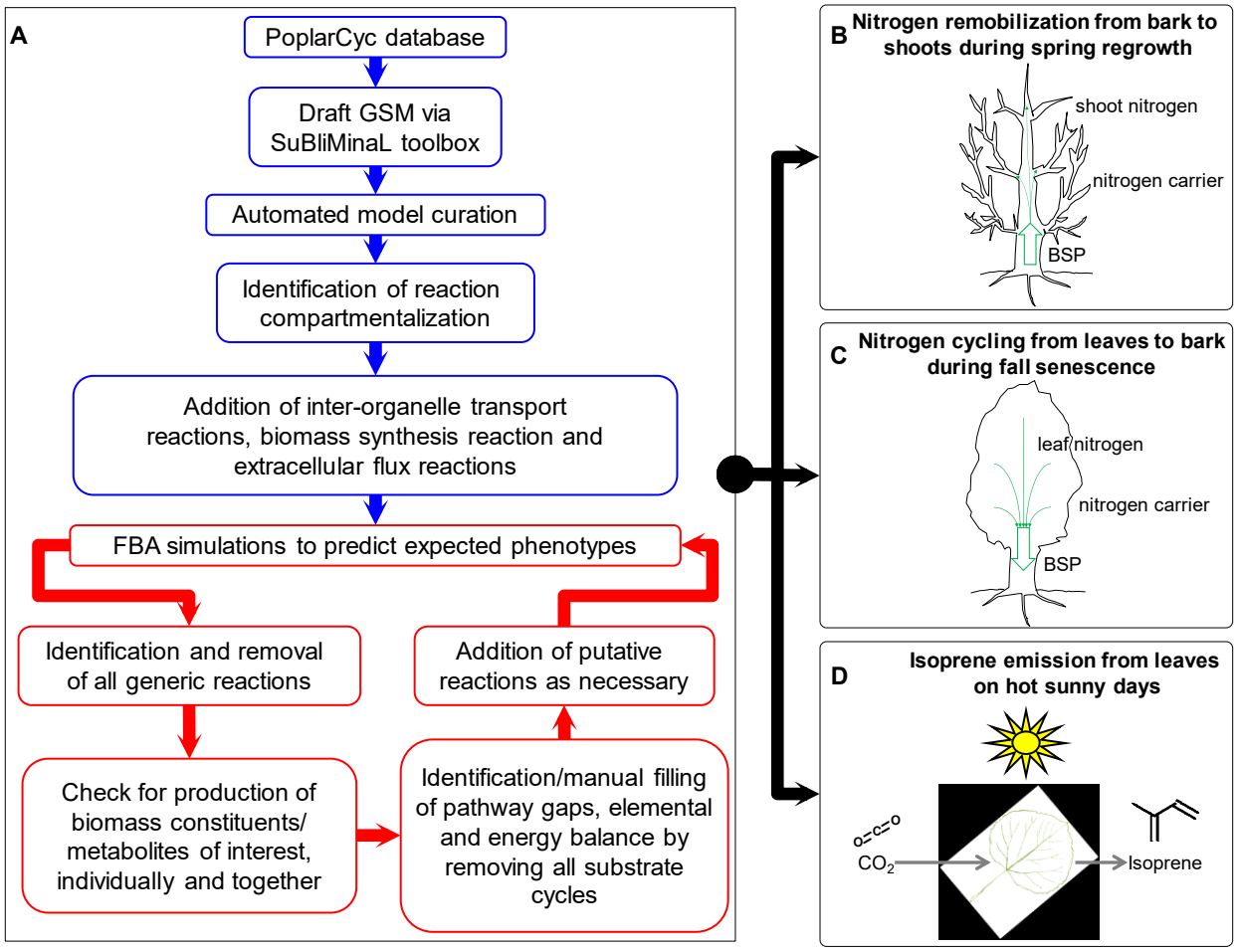


Figure 3.1. Workflow of PoplarGSM1.0 reconstruction and analysis. (A) We drafted a GSM by using the automated toolbox SuBliMinaL and curated this draft by a mixture of automated and manual methods. Red boxes and lines indicate iterative steps. We employed this model for FBA of the following three tree-specific metabolic scenarios. (B) Nitrogen remobilization from BSP to shoots during leaf regrowth in spring. (C) Nitrogen cycling and storage as BSP during leaf senescence in fall. (D) Isoprene emission by leaves. Abbreviations: BSP, bark storage protein; CO_2, carbon dioxide; GSM, genome-scale stoichiometric model.

3.3 Results

We employed the COBRA toolbox in a MATLAB environment with the open source Gnu Linear Programming Kit to solve the FBA problems in this study. The first set of flux balance analyses were performed to iteratively curate the model so that it could reasonably simulate three types of experimental observations: (i) metabolic fluxes in heterotrophic poplar cell suspensions, (ii) nitrogen mobilization in during leaf senescence and regrowth, and (iii) isoprene emission. This article describes the analysis of the last two tree-specific metabolic scenarios **(Figure 3.1.)**.

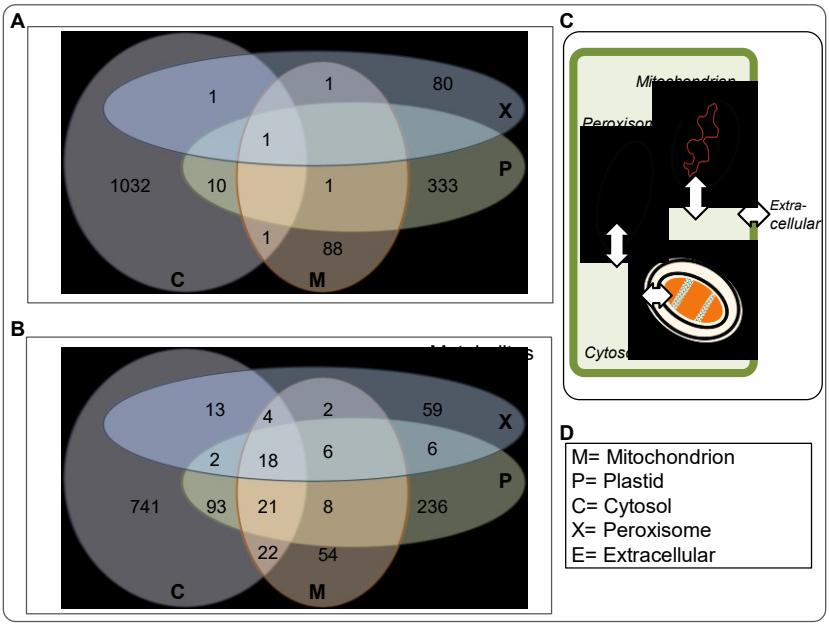


Figure 3.2. Organellar distributions of unique metabolic reactions and metabolites in PoplarGSM1.0. For each metabolic reaction in our model, we first identified the Arabidopsis gene ortholog corresponding to the gene encoding the reaction. Next, we obtained information on the organellar location of the Arabidopsis ortholog from multiple *in silico* techniques and from the literature. We then manually assigned each reaction to a compartment by (i) a *majority decision* rule (default), (ii) a *split decision* rule when different *in silico* techniques made contradictory predictions, and (iii) using compartmentalization information from the literature if none of the *in-silico* techniques positively identified an organellar location. Finally, we assigned the reaction to cytosol if all the previous strategies failed to assign an organelle. (A) Euler diagram showing distribution of reactions among organelles in PoplarGSM1.0; (B) Euler diagram showing distribution of metabolites among organelles in PoplarGSM1.0; (C) Schematic diagram of a poplar cell with the organelles in PoplarGSM1.0; (D) Abbreviations for organelles. We first performed FBA to predict biomass synthesis on glucose as the predominant carbon source in heterotrophic poplar cell suspensions, for which we have previously reported a flux model (Nargund et al., 2014; Zhang et al., 2018). Early drafts of the model were unable to simulate biomass synthesis primarily due to gaps in the metabolic network (data not shown), which required extensive curation. Because this curation process is time-consuming and impossible to carry out *ad infinitum*, we restricted the curation only to successful prediction of cellular metabolism in our cases of interest. Nevertheless, we believe that our model can readily be used to predict cellular metabolism in other tissues of interest by appropriately selecting additional inter-organelle transport reactions.

3.3.1 Nitrogen storage in bark storage protein and its remobilization during spring

We employed PoplarGSM1.0 to analyze seasonal nitrogen recycling in poplar (Cooke and Weih, 2005). This phenomenon is unique to perennial trees and occurs during the fall

and spring seasons (**Figure 3.3**) and consists of two stages: (i) cycling and (ii) remobilization. In nitrogen cycling, which occurs in fall, trees break down nitrogen from protein in senescing tissues (leaves) and relocate the nitrogen to perennial tissues (bark). The bark stores the nitrogen over winter in the form of bark storage protein (BSP). In nitrogen remobilization, which occurs during spring, trees break down the stored BSP and transport it to growing shoots (Cooke and Weih, 2005). In poplar and other trees, a nontrivial and unexplained experimental observation is the predominance of Gln as nitrogen carrier during both fall and spring (Babst and Coleman, 2018). For example, the amino acid pool in the xylem sap contains 78.3% Gln in late April (Sauter and van Cleve, 1992), which sharply contrasts with the low proportion of Gln in BSP (3.04%) (Coleman et al., 1992). The cyclic process of converting amino acids from leaf proteins to a pool of mostly Gln and then reconverting them to the 20 amino acids in BSP appears wasteful at first glance. Thus far, no studies have hypothesized or determined any metabolic benefit to poplar from this preferential conversion to Gln. Here, we use FBA to show that Gln offers an energetic advantage for regrowth of poplar shoots in early spring.

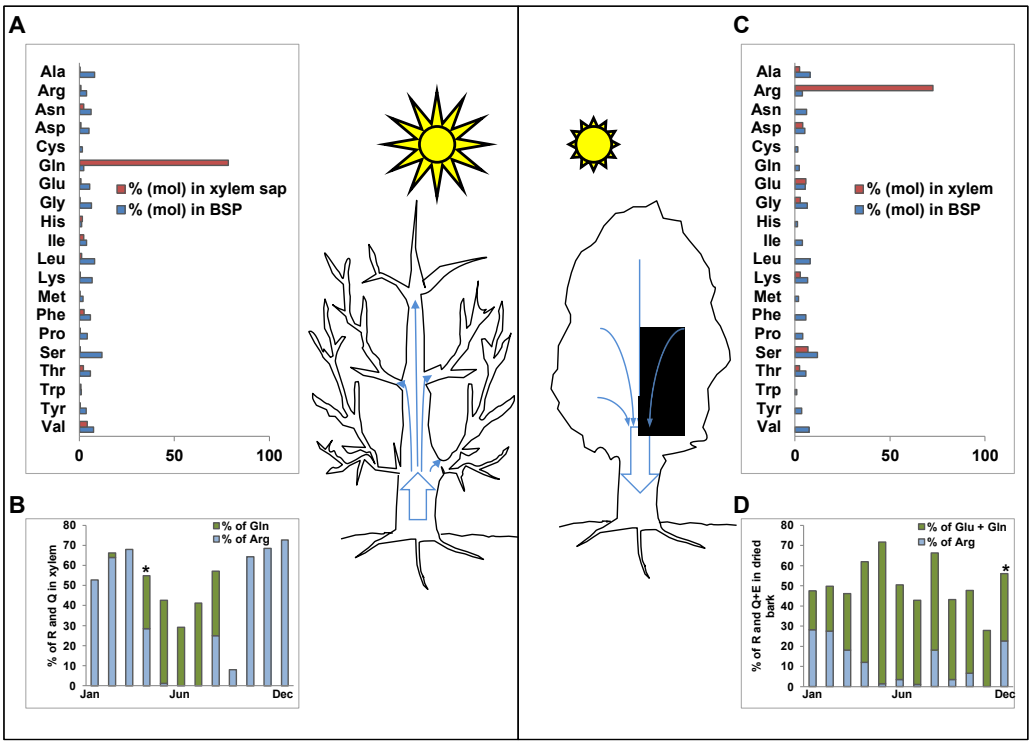


Figure 3.3. Concentration and percentage of amino acids in poplar bark and xylem. Woody plants (trees) such as poplar store nitrogen in the bark in the form of bark storage protein (BSP) during winter. In spring, they break down this protein to regenerate nitrogen for shoot growth. Experimental observations of amino acid compositions in poplar xylem sap (Sagisaka, *Plant Physiol.* 53: [319-322], 1974; Sauter and van Cleve, *Trees* 7: [26-32], 1992) show that nitrogen resulting from BSP breakdown is carried from the bark to the shoot almost exclusively in the form of glutamine, while during late winter an arginine spike is observed in the bark of poplar. However, it has been unclear if these phenomena confer any metabolic benefit to poplar. The blue arrows in the sketches show the direction of nitrogen flow, from leaf to bark during leaf senescence in fall and in

winter, and from bark to shoot in spring. A, percentage of all proteinogenic amino acids in xylem sap and bark storage protein of poplar. (Sauter and van Cleve, *Trees* 7: [26-32], 1992; Coleman *et al.*, *Plant Physiol.* 98(2): [687-693], 1992) in spring. The composition of xylem sap is for the month of May. B, contributions of Arg and Gln to the total amino acid pool in the xylem tissue throughout the year (Sagisaka, *Plant Physiol.* 53: [319-322], 1974). The maximum observed concentration of Gln in xylem sap (shown in panel A) in late April, marked with an asterisk here. C, percentage of all proteinogenic amino acids in living bark tissue and bark storage protein of poplar. (Sagisaka, *Plant Physiol.* 53: [319-322], 1974; Coleman *et al.*, *Plant Physiol.* 98(2): [687-693], 1992) in winter. The percent composition in living bark is for the month of December. D, a comparison of contributions of Arg vs Glu and Gln to the total amino acid pool in dried bark tissue (Sagisaka, *Plant Physiol.* 53: [319-322], 1974). All abbreviations of metabolites are defined in the List of Abbreviations. The month of December witnessed the highest concentration of Arg in the bark, marked with an asterisk here.

3.3.1.1 FBA predicts energetic optimality to drive the conversion of proteinogenic amino acids in BSP to Gln during nitrogen remobilization

We used FBA with a multi-tissue model consisting of interacting GSMs of bark and shoots to analyze BSP mobilization during spring. To mimic the process of BSP mobilization, FBA with the multi-tissue model involved the following: (i) all proteinogenic amino acids resulting from proteolysis of one mole of BSP in bark were converted to only one proteinogenic amino acid at a time (Figure 3.4A); (ii) the single amino acid pool was transported to shoots; (iii) the single amino acid pool was converted to biomass. The conversions to specific amino acids, and then to leaf biomass, resulted in excess nitrogen due to differences in carbon to nitrogen ratio. This excess nitrogen was permitted to be excreted out of the system as NH_3 and henceforth referred to as nitrogen loss. We constrained our model with BSP as the sole nitrogen source for biomass synthesis to reflect the environmental constraint of unsuitable soil conditions for nitrogen uptake during early spring. We permitted unrestricted exchange of all other inorganic nutrients, protons, O_2 and H_2O due to unavailability of compositional data, to the best of our knowledge, during spring BSP mobilization. Early in spring, Poplar trees cannot generate the energy necessary for biomass synthesis due to a lack of leaves. By mid-spring the poplar trees regain some photosynthetic capacity after growth of some buds and leaves. Hence, we restricted CO_2 uptake to avoid excessive photosynthesis by ribulose-bisphosphate carboxylase (RuBisCO). We also allowed additional reductant and ATP generation in shoots, if required for biomass synthesis, through photon uptake. To mimic enzymatic parsimony, we employed the objective function of biomass maximization in shoots, with L1 norm minimization, in these analyses. L1 norm minimization eliminated redundant cycles in flux solutions, a common artifact in FBA with large models. With these biologically relevant constraints, mobilization of stored nitrogen in BSP is limited only by the structure of the metabolic network. The details of the constraints employed, and their metabolic justifications are provided in Table 1.

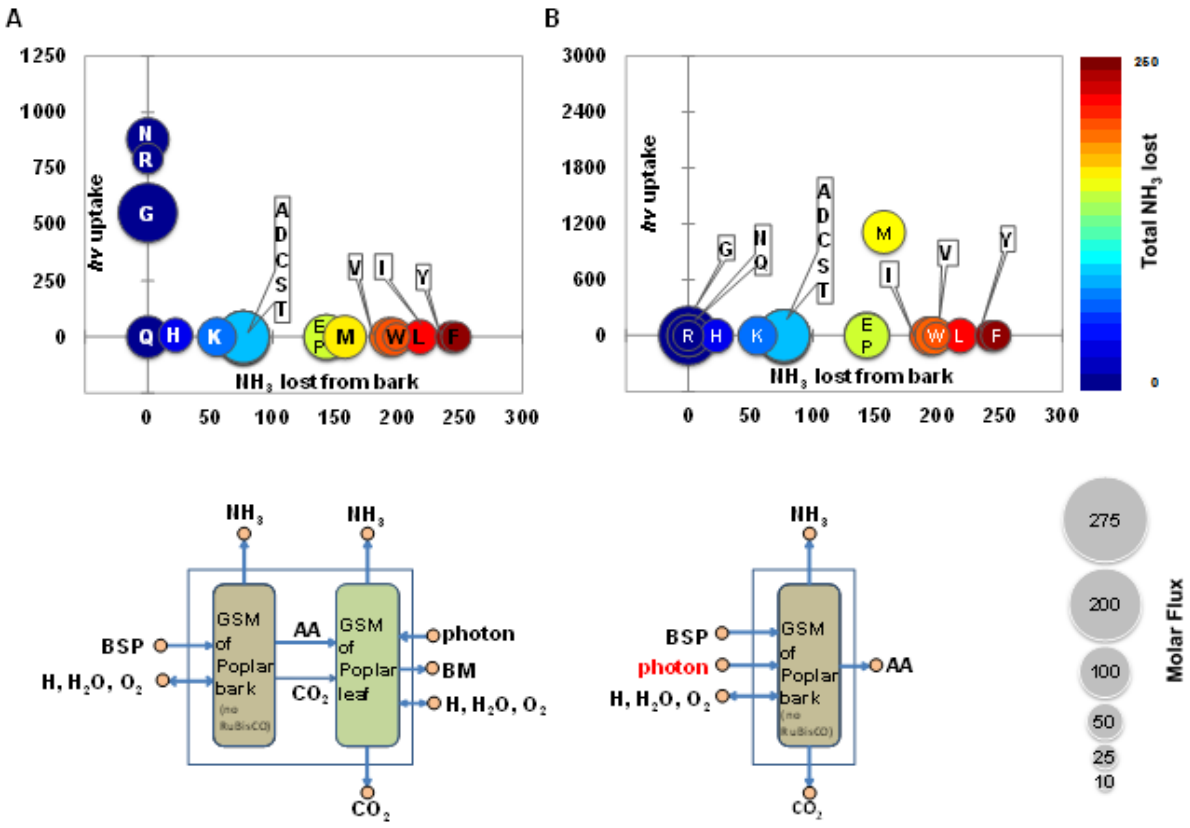


Figure 3.4. Flux balance analysis (FBA) on a poplar genome-scale metabolic model predicts glutamine to be the most energy- and nitrogen-conserving amino acid for transport of nitrogen from bark to shoot in spring, consistent with field observations. A, comparison of stored nitrogen mobilization efficiency of amino acids as predicted by a multicellular genome-scale model of poplar. Each circle represents an FBA simulation in which (i) 1 mol of the 32 kDa BSP is converted to a single amino acid whose one-letter abbreviation appears in the circle (circles with more than one letter indicate identical results for the corresponding amino acids); (ii) this amino acid is transported through the xylem to growing shoot buds; and (iii) the nitrogen (and carbon) in the amino acid is metabolized to form nitrogen-containing biomass constituents (proteins and nucleic acids) of the shoot. The plot depicts the following metrics calculated by FBA for the process described in (i) to (iii): the number of moles of nitrogen lost as NH_3 in the bark (x-axis), the minimum photon uptake required to generate the reductant and ATP necessary for the breakdown (y-axis), the total number of moles of nitrogen lost as NH_3 in the bark and shoot (circle color) and the molar flux of amino acids through xylem (circle area). FBA used a multicellular genome-scale model of poplar and employed maximization of biomass with L_1 norm minimization as its objective function, with the constraint that no nitrogen and carbon other than the pool stored in BSP was available. B, comparison of efficiencies of conversion of BSP to amino acids as predicted by a single cell genome-scale model of poplar. Each circle represents an FBA simulation in which 1 mol of the 32 kDa BSP is converted to a single

amino acid whose one-letter abbreviation appears in the circle (circles with more than one letter indicate identical results for the corresponding amino acids. The plot depicts the following metrics calculated by FBA: the number of moles of nitrogen lost as NH_3 in the bark (*x*-axis and circle color), the minimum photon uptake required to generate the reductant and ATP necessary for the breakdown (*y*-axis) and the molar flux of amino acids through xylem (circle area). FBA used a single cell genome-scale model of poplar and employed maximization of amino acids with L_1 norm minimization as its objective function, with the constraint that no nitrogen and carbon other than the pool stored in BSP was available. All abbreviations of metabolites are defined in the List of Abbreviations. The sketches below the plots represent the constraints related to C and N flow as employed in the simulations. The details and metabolic justifications of the constraints on the simulations of **Figure 3.4**A and **Figure 3.4**B are included in Table 1 and **Table 2** respectively.

The analyses predicted that conversion of BSP to four amino acids (Arg, Asn, Gln and Gly) could mobilize the entire nitrogen pool without any nitrogen loss in bark or shoot **(Figure 3.4A)**. However, only in case of Gln, it did not require generation of additional reductant or ATP through photon uptake **(Figure 3.4A)**. Thus, Gln is the optimal nitrogen transport amino acid once poplar trees gain some photosynthetic capacity. However, the percentage of Gln in xylem sap during early spring, when poplar lacks the photosynthetic machinery, is significantly less when compared to late spring, despite being the predominant amino acid (Sauter and van Cleve, 1992).

We uncoupled bark metabolism from shoots metabolism by performing a series of FBA simulations with our unicellular PoplarGSM1.0, One mole of BSP was converted to one amino acid at a time **(Figure 3.4B)** in bark. We allowed a hypothetical photon uptake by bark to generate reductant and ATP, to account for reductant and ATP generation by breakdown of stored carbohydrates in bark. The simulations predicted that BSP could be mobilized without any nitrogen loss to only four amino acids (Arg, Asn, Gln, and Gly) without any additional energetic need, the same set predicted in the multi-tissue FBA simulations of early spring BSP mobilization. However, large quantities of inorganic nitrogen in the bark or the shoots has not been observed in field studies. Additionally, xylem sap contains all twenty proteinogenic amino acids (Sauter and van Cleve, 1992). These predictions, in conjunction with multi-tissue simulations of early spring mobilization, indicate that until poplar attains the photosynthetic machinery to fix CO_2, a preferential conversion to Gln minimizes nitrogen loss. The breakdown of stored carbohydrate reserve, which has been observed in field studies, could also play a role in minimizing nitrogen loss. The details of the constraints employed, and their metabolic justifications are provided in **Table 2**.

Table 1. Constraints on multi-tissue FBA for BSP mobilization in spring

This table displays the constraints placed on the bark and the shoots in the multi-tissue FBA simulations for late spring BSP mobilization. The constraints placed on the tissues are placed to predict the N transport efficiency of amino acids from the shoots to the bark during BSP mobilization in spring.

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{bark} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during BSP mobilization
$v^{bark}_{RuBisCo} = 0$	Blocked RuBisCo activity	RuBisCo does not fix carbon in bark
$-1000 \leq \quad \leq$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$v^{bark}_{BSP} = -1$	1 mol of BSP uptake	BSP is the sole source of C and N to predict parsimonious nutrient use
$0 \leq v^{bark}_{NH_3} \leq 1000$	Permitted N loss as NH_3	N in BSP not transported to shoots as amino acids is excreted as NH_3
$0 \leq v^{bark}_{CO_2} \leq 1000$	Permitted CO_2 excretion	C in BSP not assimilated in amino acid is excreted as CO_2
$-1000 \leq v^{bark}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
Constraints on transport	**Equations and Bounds**	**Justification**
$0 \leq v^{bark\ to\ leaf}_{amino\ acid} \leq 1000$	N transport from bark to shoot	N is transported to shoots exclusively as amino acid, one at a time
$0 \leq v^{bark\ to\ leaf}_{CO_2} \leq 1000$	CO_2 transport from bark to shoots	C not transported as amino acid can be transported to shoots as CO_2
Constraints on shoots	**Equations and Bounds**	**Justification**
$\mathbf{S}.\mathbf{v}^{leaf} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during BSP mobilization
$0 \leq v^{leaf}_{RuBisCo} \leq 1000$	Active RuBisCo in green shoots	RuBisCo in green shoots can fix CO_2 with sufficient photoperiod in spring
$-1000 \leq v^{leaf}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{leaf}_{NH_3} \leq 1000$	Permitted N loss as NH_3	Transported N not assimilated in biomass in shoots is presumed lost
$-1000 \leq v^{leaf}_{CO_2} \leq 0$	Permitted CO_2 uptake	Shoots with active RuBisCo can photosynthetically fix additional CO_2
$-1000 \leq v^{leaf}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
$-8000 \leq v^{leaf}_{h} \leq 0$	Permitted photon uptake	RuBisCo can capture photons during spring to carry out photosynthesis

Table 2. Constraints on single tissue FBA for BSP mobilization in the bark tissue spring

This table displays the constraints placed on the bark tissue in the single-tissue FBA simulations for late spring BSP mobilization. The constraints placed on the tissues are placed to predict the conversion efficiency of BSP to amino acids while preserving N .

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{\mathrm{bark}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during BSP mobilization
$\mathrm{v}^{\mathrm{bark}}_{\mathrm{RuBisCo}} = 0$	Blocked $\mathrm{RuBisCo}$ activity	$\mathrm{RuBisCo}$ does not fix carbon in bark
$-1000 \leq \mathrm{v}^{\mathrm{bark}}_{\mathrm{H,O_2,H_2O}} \leq 1000$	Unrestricted exchange of $\mathrm{H}, \mathrm{O}_2, \mathrm{H}_2\mathrm{O}$	No reported data on $\mathrm{H}, \mathrm{O}_2, \mathrm{H}_2\mathrm{O}$ exchange fluxes
$\mathrm{v}^{\mathrm{bark}}_{\mathrm{BSP}} = -1$	1 mol of BSP uptake	BSP the sole source of C and N to predict parsimonious nutrient use
$0 \leq \mathrm{v}^{\mathrm{bark}}_{\mathrm{NH_3}} \leq 1000$	Permitted N loss as NH_3	N in BSP not assimilated in amino acids is excreted as NH_3
$0 \leq \mathrm{v}^{\mathrm{bark}}_{\mathrm{CO_2}} \leq 1000$	Permitted CO_2 excretion	C in BSP not assimilated in amino acids excreted from bark as CO_2
$-1000 \leq \mathrm{v}^{\mathrm{bark}}_{\mathrm{Pi,SO_4}} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
$0 \leq \mathrm{v}^{\mathrm{bark\ to\ leaf}}_{\mathrm{amino\ acid}} \leq 1000$	N transport from bark to shoot	N transported to shoots exclusively as amino acid, one at a time
$-8000 \leq \mathrm{v}^{\mathrm{bark}}_{\mathrm{h}} \leq 0$	Permitted photon uptake	Hypothetical photon uptake in bark a measure of additional energetic need

The previous simulations could not explain the gradual increase in Gln content in xylem sap as spring progressed. To explain it we repeated the multi-tissue FBA simulations after eliminating RuBisCO activity and photon uptake, with all other constraints remaining the same **(Figure 3.5)**. In this case of BSP mobilization in early spring, FBA predicted that Arg, Asn, Gln and Gly could transport the nitrogen pool without any nitrogen loss in bark. However, without RuBisCO activity in the shoots the entire transported nitrogen pool in these four amino acids could not be assimilated in biomass, leading to N loss in the shoots. On the contrary, in the early spring simulations, conversion of BSP to multiple amino acids (Ala, Asp, Cys, Ser, Thr, Glu, Pro, Val, Ile, Leu, Tyr, Arg, Phe, and Trp) led to nitrogen loss in bark. However, these fourteen amino acids could be metabolized to biomass without any nitrogen loss in the shoots. Without photon uptake and RuBisCO activity, no amino acid could transport the entire pool of stored nitrogen in BSP and then undergo metabolism to synthesize biomass. In terms of total nitrogen loss in the bark and the shoots, Ala, Asp, Cys, Ser and Thr were the most efficient amino acids, followed by Gln. Thus, without RuBisCO activity in the shoots during early spring, a decrease in the predominance of Gln in the xylem sap optimizes the utilization of stored nitrogen reserve.

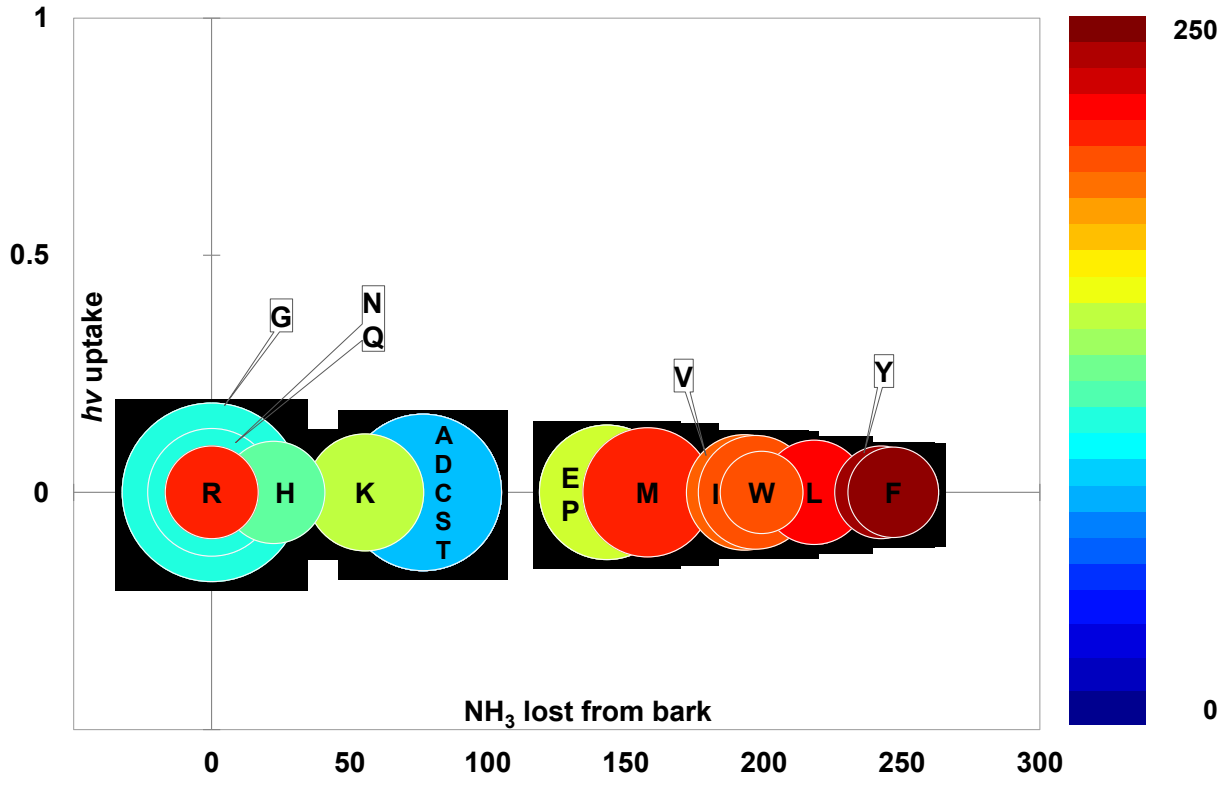


Figure 3.5. Flux balance analysis (FBA) on a poplar genome-scale metabolic model, with zero RuBisCO activity in the shoots, predicts multiple amino acids to be equally efficient for transport of nitrogen from bark to shoot in spring, consistent with field observations. A, comparison of stored nitrogen mobilization efficiency of amino acids as predicted by a multicellular genome-scale model of poplar. Each circle represents an FBA simulation in which (i) 1 mol of the 32 kDa BSP is converted to a

single amino acid whose one-letter abbreviation appears in the circle (circles with more than one letter indicate identical results for the corresponding amino acids); (ii) this amino acid is transported through the xylem to growing shoot buds; and (iii) the nitrogen (and carbon) in the amino acid is metabolized to form nitrogen-containing biomass constituents (proteins and nucleic acids) of the shoot. The plot depicts the following metrics calculated by FBA for the process described in (i) to (iii): the number of moles of nitrogen lost as NH_3 in the bark (*x*-axis), the minimum photon uptake required to generate the reductant and ATP necessary for the breakdown (*y*-axis), the total number of moles of nitrogen lost as NH_3 in the bark and shoot (circle color) and the molar flux of amino acids through xylem (circle area). FBA used a multicellular genome-scale model of poplar and employed maximization of biomass with L_1 norm minimization as its objective function, with the constraint that no nitrogen and carbon other than the pool stored in BSP was available.

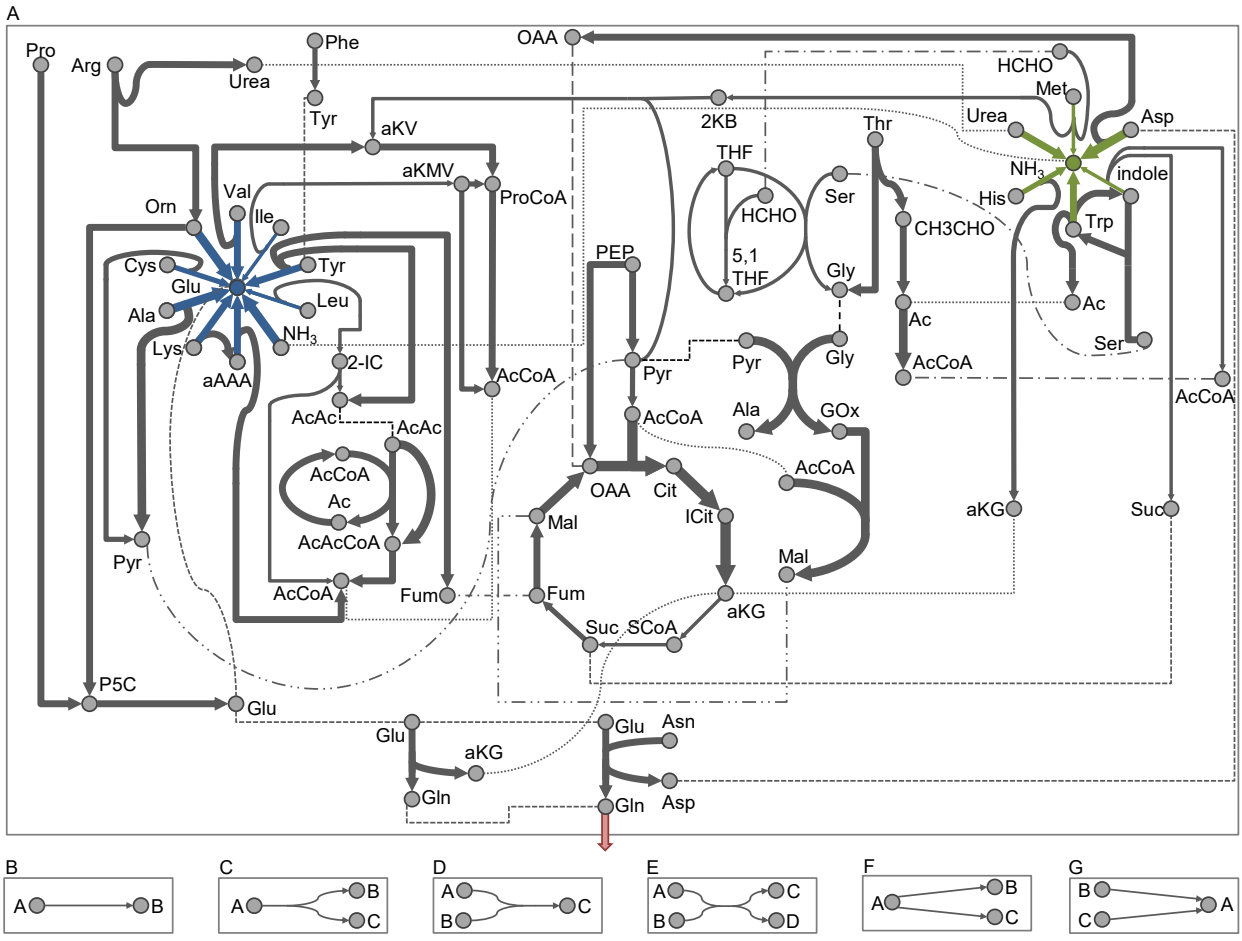


Figure 3.6. Optimal route for early spring BSP mobilization as predicted by FBA. A, flux map displaying the reactions in the bark of poplar during mobilization BSP to only Gln. In an FBA simulation, 1 mole of BSP is converted to Gln in poplar bark; Gln is then transported to the shoots and metabolized there to biomass. The FBA simulation was performed using the multicellular compartmented GSM and employed the objective

function of maximization of biomass with L_1 norm minimization, with the constraints of no additional N or C other than the pool stored in biomass. To aid visualization, the compartmented FBA solution was converted to a simplified uncompartmented solution, which was then used to prepare the flux map. The conversion involved three major steps, (i) removal of compartmentation from the reactions; (ii) removal of cofactors from the reactions, with the objective of observing only the C and N flow; (iii) addition of fluxes in order to calculate a total overall flux when a reaction was predicted to take place in multiple compartments. To utilize space better in the figure, multiple reactions carrying the same flux were often combined to a single overall reaction. CO_2 is treated as a cofactor, and in case of either assimilation or release of CO_2 in a reaction in a pathway, only the initial and final metabolites are shown. Each circle in the flux map represents a metabolite and each solid arrow represents a metabolic reaction or pathway. The dashed lines are used as connectors to connect a metabolite appearing multiple times in the flux map, whenever possible. The single red arrow crossing the solid boundary represents the transport of Gln from the bark to the shoots. The arrow thicknesses represent the flux through that reaction in a logarithmic scale. The blue circle represents Glu as the nitrogen carrier in which the metabolites connected to it by solid arrows donate their N to aKG (not shown in the map), thereby forming Glu. Similarly, the green circle represents NH_3 released in degradation pathways by the metabolites connected to it by solid arrows. A description of the schematics used to represent reactions or pathways, as used in the preparation of the flux map, is provided in panels B-G. All abbreviations of metabolites are defined in the List of Abbreviations. B, the schematic representation of a uni-uni (unireactant and uniproduct) reaction. C, the schematic representation of a uni-bi (unireactant and biproduct) reaction. D, the schematic representation of a bi-uni (bireactant and uniproduct) reaction. E, the schematic representation of a bi-bi (bireactant and biproduct) reaction. F, the schematic representation of a reactant getting converted to two products by two different reactions or pathways reaction. G, the schematic representation of two reactants getting converted to a product by two different reactions or pathways.

We prepared a flux map **(**
Figure 3.6) of preferential BSP mobilization to only Gln in early spring conditions. FBA predicted significant flux through the amino acid degradation pathways, expected due to the constraints on the model. The glyoxylate cycle operated with significant flux from malate synthase to aconitase. Additionally, significant flux through acetoacetyl-CoA synthase and the anaplerotic reaction of PEP carboxylase were. Furthermore, FBA predicted significant turnover of acetoacetyl-CoA to acetyl-CoA. The TCA cycle was fueled by acetyl-CoA and the anaplerotic reaction. These predictions are consistent with experimentally observed increase in transcript levels of genes associated with enzymes catalyzing these reactions (Druart et al., 2007).

3.3.1.2 Leaf nitrogen mobilization during leaf senescence as Gln was predicted suboptimal

We used FBA with the multi-tissue model to analyze nitrogen mobilization from senescing leaves during fall. In addition to the leaf nitrogen content, significant amount of sucrose (Suc) is also transported to bark for storage during fall. To mimic this process

of nutrient relocation, FBA with the multi-tissue model involved the following: (i) all biomass constituents resulting from degradation of one mole of biomass were converted to only one proteinogenic amino acid, and a specific amount of Suc, at a time; (ii) the pool consisting of the amino acid and Suc was transported to bark; (iii) the transported nutrient pool was converted to all other proteinogenic amino acids, which then constitute BSP. The conversions of the transported nutrient pool to BSP, resulted in excess nitrogen. Excretion of excess nitrogen out of the system as NH_3 was permitted. Additionally, we permitted unrestricted exchange of all other inorganic nutrients, protons, O_2 and H_2O due to unavailability of compositional data. To mimic enzymatic parsimony, and to eliminate redundant cycles in the solution, we employed the objective function of biomass maximization in shoots with L_1 norm minimization, in these analyses.

Due to differences in carbon-to-nitrogen ratios of BSP and the transported metabolites, CO_2 generation was predicted in poplar bark during the formation of BSP. Amount of CO_2 generation varied depending upon the amount of transported Suc which was specified as a fraction of total C in leaf biomass. It is unclear is any CO_2 transport from bark to leaf occurs during fall. Additionally, it is unclear if photosynthesis stops completely during gradual RuBisCO breakdown in fall. Hence, we considered all these scenarios and performed simulations in which we permitted: (a) photosynthesis without any CO_2 exchange between bark and leaves **(Figure 3.7)**; (b) transport of CO_2 from bark to leaves and photon uptake without additional CO_2 uptake in leaves **(Figure 3.8)**; (c) photon uptake without any CO_2 uptake in leaves or mixing of CO_2 pools released in bark and leaves **(Figure 3.9)**; (d) no photon uptake and no CO_2 exchange between bark and leaves **(Figure 3.10)**.

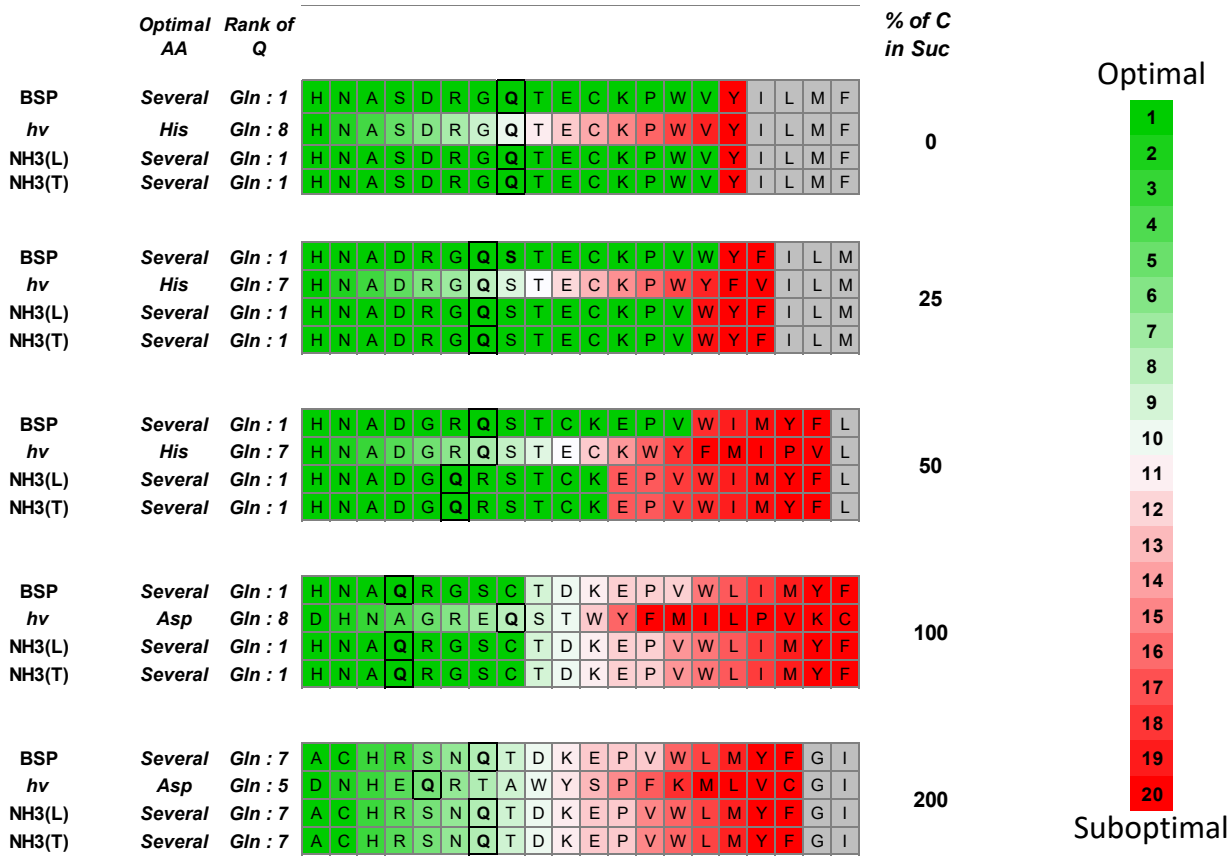


Figure 3.7. Flux balance analysis (FBA) on a poplar genome-scale metabolic model predicts glutamine to be suboptimal, when CO_2 uptake and photosynthesis are permitted, for during leaf nitrogen mobilization in fall. A comparison of nitrogen mobilization efficiency of all proteinogenic amino acids as predicted by a multicellular genome-scale model of poplar, under varying metabolic demand of BSP synthesis with sucrose accumulation. The five grids in the figure depict the five series of simulations with different amounts of sucrose accumulation with fall leaf N mobilization, with the amount indicated to the right of the grid. The percentage of sucrose synthesized is calculated based on the number of moles of C atoms contained in the leaf biomass. The demand of sucrose biosynthesis in the leaves is gradually increased from 0% to 200% to account for varying rates of Suc accumulation during fall. Each square in a grid represents an FBA simulation in which (i) the nitrogen containing constituents in 1 mol of biomass is converted to the amino acid whose one-letter abbreviation appears in the square, with the additional metabolic demand of diversion of the specified minimum amount of C to sucrose biosynthesis and storage (ii) the amino acid and the sucrose are transported to bark; and (iii) the amino acid and sucrose are metabolized to form the 32-kDA bark storage protein. The CO_2 released in either leaf or bark during this process cannot be transported to the other tissue. FBA used a multicellular genome-scale model of poplar and employed maximization of BSP with L_1 norm minimization as its objective function and the constraint that no nitrogen other than the pool stored in biomass is available. Optimality of the amino acids are evaluated in terms of four biologically relevant metrics, (i) the amount of BSP synthesized in the bark; (ii) the amount of photon uptake; (iii) the number of moles of nitrogen lost as NH_3 in the leaf and (iv) the total

number of moles of nitrogen lost as NH_3 in the leaf and bark. The amino acids are arranged in a row in the decreasing order of optimality, based on each metric, within a grid for each series of simulations. The optimal amino acids and the rank of Gln for each series of simulations are mentioned for each metric, with the position of Gln placed in a box with thick borders. The grey boxed within a grid indicates simulations that did not converge. The details and metabolic justifications of the constraints on these simulations are included in **Table 3.** .

We ranked the amino acids on four metrics of amount of BSP formed, photon uptake, NH_3 lost in the leaf and total NH_3 lost (in the bark and leaf) for each series of simulations. An amino acid optimal on all four metrics is the ideal choice in N transport from the senescing leaves to bark. The analyses predicted Gln to be suboptimal nitrogen transport amino acid based on all four metrics, regardless of constraints on photosynthesis or mixing of CO_2 pools **(Figure 3.7, Figure 3.8, Figure 3.9 and Figure 3.10)**. However, Gln was equally optimal as multiple other amino acids on multiple metrics. Additionally, the optimal amino acid varied depending upon the constraints imposed during the simulations. Additionally, within each series of simulation, the optimal amino acid evaluated on a single metric were not always identical. The optimality rank of the amino acids changed with the amount of sucrose transported from the leaves to the bark, all other constraints remaining the same.

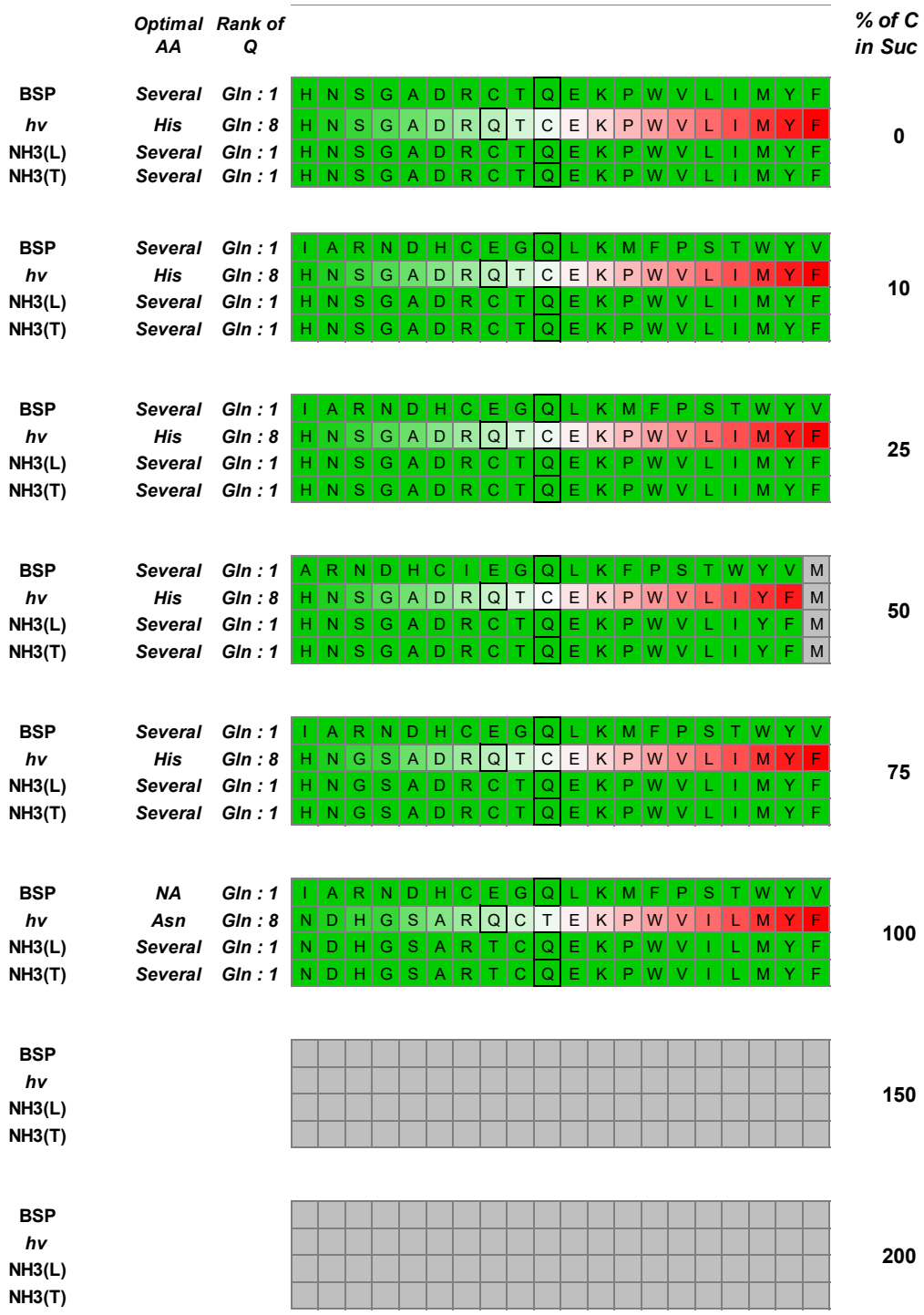


Figure 3.8. Flux balance analysis (FBA) on a poplar genome-scale metabolic model predicts glutamine to be suboptimal on optimal reductant utilization during leaf nitrogen mobilization in fall, with transport of CO_2 pool released in bark to senescing leaves and photon uptake without additional CO_2 uptake in leaves. A comparison of nitrogen mobilization efficiency of all proteinogenic amino acids as predicted by a multicellular genome-scale model of poplar, under varying metabolic demand of BSP synthesis with sucrose accumulation. The eight grids in the figure depict

the eight series of simulations with different minimum amount of sucrose accumulation, with the amount indicated to the right of the grid. The percentage of sucrose synthesized is calculated based on the number of moles of C atoms contained in the leaf biomass. The demand of sucrose biosynthesis in the leaves is gradually increased from 0% to 200% to account for varying rates of Suc accumulation during fall. Each square in a grid represents an FBA simulation in which (i) the nitrogen containing constituents in 1 mol of biomass is converted to the amino acid whose one-letter abbreviation appears in the square, with the additional metabolic demand of diversion of the specified minimum amount of C to sucrose biosynthesis and storage (ii) the amino acid and the sucrose are transported through the xylem to bark; and (iii) the amino acid and sucrose are metabolized to form the 32-kDA bark storage protein. The CO_2 released in bark during this process can either be transported to the leaf for fixation, depending upon availability of reductant, or released. FBA used a multicellular genome-scale model of poplar and employed maximization of BSP with L_1 norm minimization as its objective function with the constraint that no nitrogen or carbon other than the pool stored in biomass is available. Optimality of the amino acids are evaluated in terms of four biologically relevant metrics, (i) the amount of BSP synthesized in the bark; (ii) the amount of photon uptake; (iii) the number of moles of nitrogen lost as NH_3 in the leaf and (iv) the total number of moles of nitrogen lost as NH_3 in the leaf and bark. The amino acids are arranged in a row in the decreasing order of optimality, based on each metric, within a grid for each series of simulations. The optimal amino acids and the rank of glutamine for each series of simulations are mentioned for each metric, with the position of Gln placed in a box with thick borders. The grey boxed within a grid indicates simulations that did not converge. The details and metabolic justifications of the constraints on these simulations are included in **Table 4**.

Table 3. Constraints on multi-tissue FBA for leaf N mobilization in fall

This table displays the constraints placed on the senescing leaves and the bark in the multi-tissue FBA simulations for leaf N mobilization. We permitted photosynthesis without any CO_2 exchange between bark and leaves. The constraints placed on the tissues are placed to predict the N transport efficiency of amino acids from the senescing leaves to the bark for storage as BSP.

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{\text{bark}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf N mobilization
$v^{\text{bark}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo activity	RuBisCo does not fix carbon in bark
$-1000 \leq v^{\text{bark}}_{\text{H,O}_2,\text{H}_2\text{O}} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{bark}}_{\text{NH}_3} \leq 1000$	Permitted N loss as NH_3	Transported N that cannot be stored in BSP is excreted as NH_3
$0 \leq v^{\text{bark}}_{\text{CO}_2} \leq 1000$	CO_2 excretion is permitted	C in transported amino acids not assimilated in BSP excreted as CO_2
$-1000 \leq v^{\text{bark}}_{\text{Pi,SO}_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
Constraints on transport	**Equations and Bounds**	**Justification**
$- \leq \leq$	N transport from leaves to bark	N is transported to the bark exclusively as amino acid, one at a time
$v^{\text{min}}_{\text{Sucrose}} \leq \leq$	Suc transport from leaves to bark	Specified minimum amount of Suc transport from leaf to bark
$v^{\text{bark to leaf}}_{\text{CO}_2} =$	CO_2 transport from bark to shoot	CO_2 generated in bark during BSP synthesis cannot be transported to leaf
Constraints on leaf	**Equations and Bounds**	**Justification**
$\mathbf{S}.\mathbf{v}^{\text{leaf}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf*N* mobilization
$v^{\text{leaf}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo in leaves	RuBisCo activity drops significantly with leaf senescence
$v^{\text{leaf}}_{\text{biomass}} = -1$	1 mol of biomass uptake	Biomass is the sole source of C and N to predict parsimonious nutrient
use $-1000 \leq v^{\text{leaf}}_{\text{H,O}_2,\text{H}_2\text{O}} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{leaf}}_{\text{NH}_3} \leq 1000$	Permitted N loss as NH_3	N in leaf that cannot be transported to bark as amino acid is excreted
$-1000 \leq v^{\text{leaf}}_{\text{CO}_2} \leq 1000$	Unrestricted CO_2 exchange	Unclear if CO_2 is assimilated or synthesized during leaf senescence
$-1000 \leq v^{\text{leaf}}_{\text{Pi,SO}_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
$v^{\text{leaf}}_{h\nu} \leq$	Photon uptake is not permitted	Photons cannot be captured in leaves once RuBisCo is broken down

Table 4. Constraints on multi-tissue FBA for leaf N mobilization in fall

This table displays the constraints placed on the senescing leaves and the bark in the multi-tissue FBA simulations for leaf N mobilization. We permitted transport of CO_2 from bark to leaves and photon uptake without additional CO_2 uptake in leaves. The constraints placed on the tissues are placed to predict the N transport efficiency of amino acids from the senescing leaves to the bark for storage as BSP.

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{\text{bark}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf N mobilization
$v^{\text{bark}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo activity	RuBisCo does not fix carbon in bark
$-1000 \leq v^{\text{bark}}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{bark}}_{NH_3} \leq 1000$	Permitted N loss as NH_3	Transported N that cannot be stored in BSP is excreted as NH_3
$0 \leq v^{\text{bark}}_{CO_2} \leq 1000$	CO_2 excretion is permitted	C in transported amino acids not assimilated in BSP excreted as CO_2
$-1000 \leq v^{\text{bark}}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
Constraints on transport	**Equations and Bounds**	**Justification**
$- \leq \leq$	N transport from leaves to bark	N is transported to the bark exclusively as amino acid, one at a time
$v^{\text{min}}_{\text{Sucrose}} \leq \leq$	Suc transport from leaves to bark	Specified minimum amount of Suc transport from leaf to bark
$0 \leq \leq$	CO_2 transport from bark to shoot	CO_2 generated in bark during BSP synthesis must be excreted
Constraints on leaf	**Equations and Bounds**	**Justification**
$\mathbf{S}.\mathbf{v}^{\text{leaf}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf*N* mobilization
$v^{\text{leaf}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo in leaves	RuBisCo activity drops significantly with leaf senescence
$v^{\text{leaf}}_{\text{biomass}} = -1$	1 mol of biomass uptake	Biomass is the sole source of C and N to predict parsimonious nutrient
use $-1000 \leq v^{\text{leaf}}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{leaf}}_{NH_3} \leq 1000$	Permitted N loss as NH_3	N in leaf that cannot be transported to bark as amino acid is excreted
$0 \leq \leq$	CO_2 excretion only from leaf	Unclear if CO_2 is assimilated or synthesized during leaf senescence
$-1000 \leq v^{\text{leaf}}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
$v^{\text{leaf}}_{h\nu} \leq$	Photon uptake is not permitted	Photons cannot be captured in leaves once RuBisCo is broken down

The multi-tissue simulations predicted that when CO_2 uptake and photosynthesis were allowed, without sucrose accumulation Gln is optimal, along with multiple other amino acids **(Figure 3.7)** on the metrics of BSP accumulation, NH_3 lost in leaves and total NH_3 lost in leaves and bark. However, photon uptake to generate reductant was necessary to utilize Gln as the N transport amino acid to mobilize the entire leaf nitrogen content. His, a positively charged amino acid, is the only amino acid that was optimal on all four metrics.

Gln stays optimal on all metrics other than photon uptake with an increase in the amount of sucrose accumulated. However, when the amount of carbon stored as sucrose increases significantly, Gln becomes suboptimal on all four metrics. When photon uptake is constrained to zero, Gln was almost always suboptimal **(Figure 3.10)**. The entire leaf nitrogen content cannot be transported for storage as BSP in the absence of photon uptake. Additionally, when CO_2 uptake is not permitted, the maximum amount of N that can be stored BSP is 64%. This is considerably less than 75-80% reported earlier. This indicates some photosynthesis does take place during leaf senescence. When CO_2 uptake and photosynthesis is permitted, the entire leaf N content can be transported as Gln and stored in BSP.

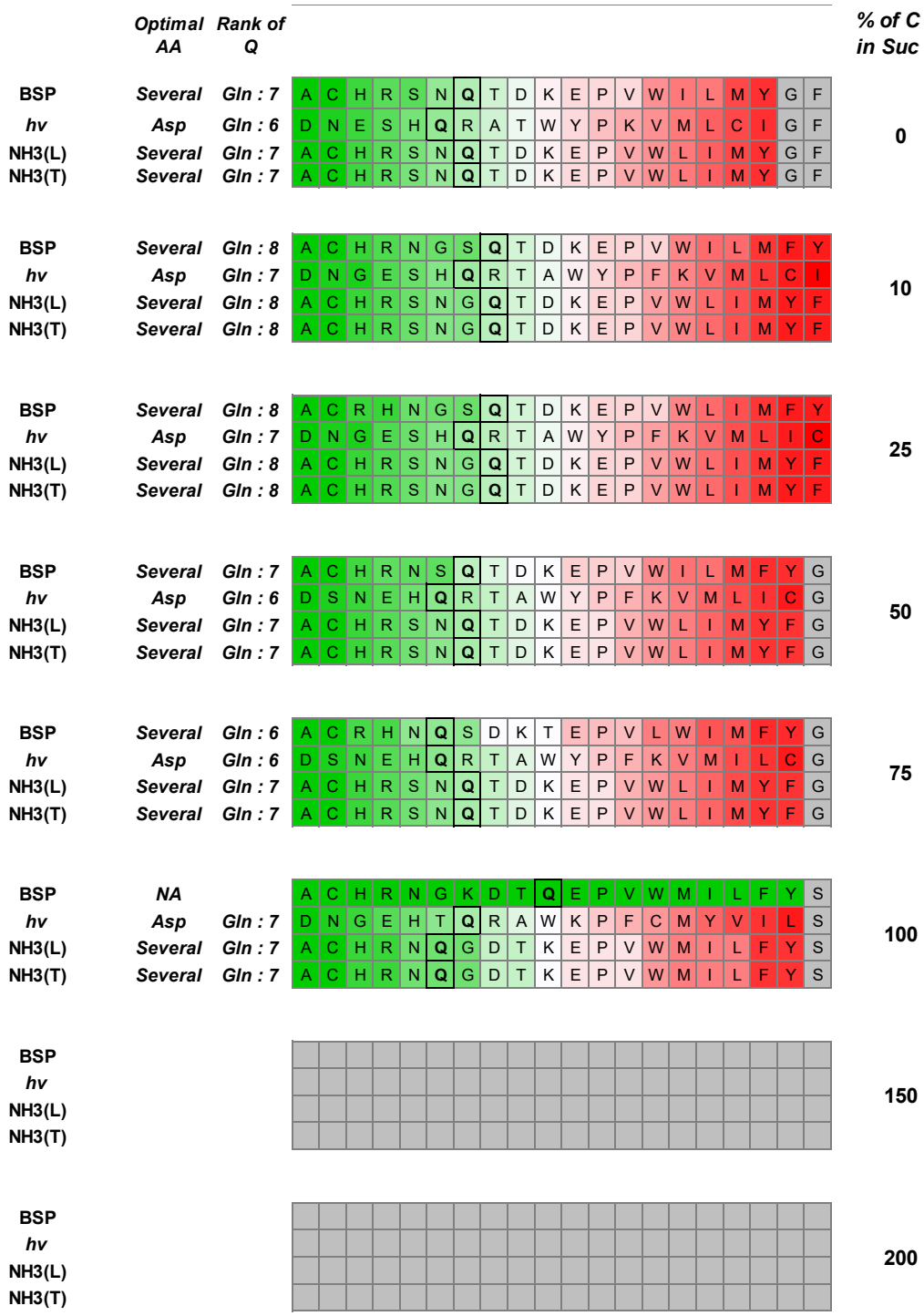


Figure 3.9. Flux balance analysis (FBA) on a poplar genome-scale metabolic model predicts glutamine to be suboptimal on multiple biologically relevant metrics during leaf nitrogen mobilization in fall, with photon uptake but without any CO_2 uptake or mixing of released CO_2 pools. A comparison of nitrogen mobilization efficiency of all proteinogenic amino acids as predicted by a multicellular genome-scale model of poplar, under varying metabolic demand of BSP synthesis with sucrose accumulation. The eight grids in the figure depict the eight series of simulations with different minimum amount of sucrose accumulation, with the amount indicated to the right of the grid. The

percentage of sucrose synthesized is calculated based on the number of moles of C atoms contained in the leaf biomass. The demand of sucrose biosynthesis in the leaves is gradually increased from 0% to 200% to account for varying rates of Suc accumulation during fall. Each square in a grid represents an FBA simulation in which (i) the nitrogen containing constituents in 1 mol of biomass is converted to the amino acid whose one-letter abbreviation appears in the square, with the additional metabolic demand of diversion of the specified minimum amount of C to sucrose biosynthesis and storage (ii) the amino acid and the sucrose are transported through the xylem to bark; and (iii) the amino acid and sucrose are metabolized to form the 32-kDA bark storage protein. The CO_2 released in leaf or bark during this process must be released. FBA used a multicellular genome-scale model of poplar and employed maximization of BSP with L_1 norm minimization as its objective function and the constraint that no nitrogen or carbon other than the pool stored in biomass is available. The optimality of the amino acids are evaluated in terms of four biologically relevant metrics, (i) the amount of BSP synthesized in the bark; (ii) the amount of photon uptake; (iii) the number of moles of nitrogen lost as NH_3 in the leaf and (iv) the total number of moles of nitrogen lost as NH_3 in the leaf and bark. The amino acids are arranged in a row in the decreasing order of optimality, based on each metric, within a grid for each series of simulations. The optimal amino acids and the rank of glutamine for each series of simulations are mentioned for each metric, with the position of Gln placed in a box with thick borders. The grey boxed within a grid indicates simulations that did not converge. The details and metabolic justifications of the constraints on these simulations are included in **Table 5**.

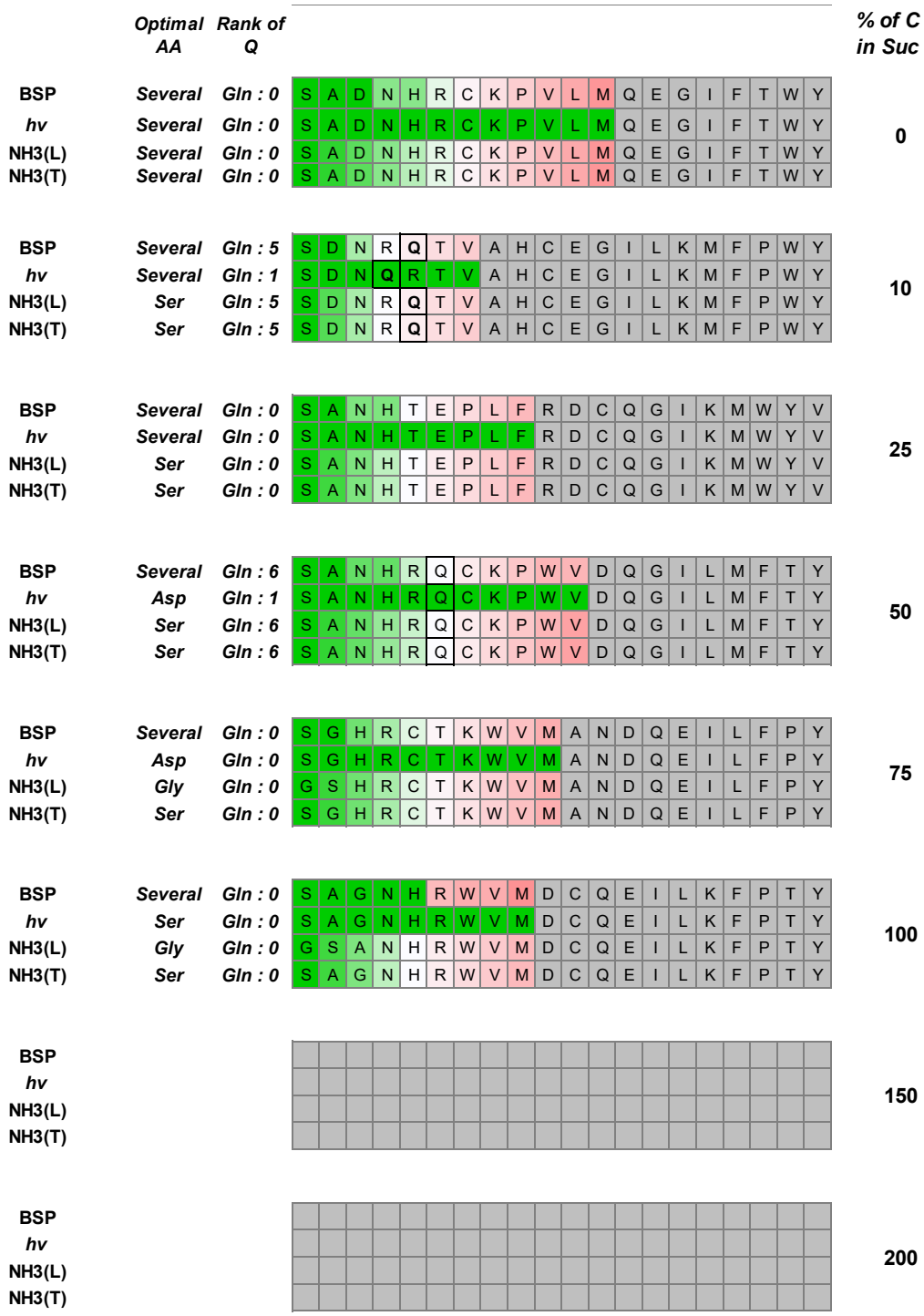


Figure 3.10. Flux balance analysis (FBA) on a poplar genome-scale metabolic model predicts glutamine to be suboptimal on multiple biologically relevant metrics during leaf nitrogen mobilization in fall, without reductant generated from photosynthesis. A comparison of nitrogen mobilization efficiency of all proteinogenic amino acids as predicted by a multicellular genome-scale model of poplar, under varying metabolic demand of BSP synthesis with sucrose accumulation. The eight grids in the figure depict the eight series of simulations with different minimum amount of sucrose accumulation,

with the amount indicated to the right of the grid. The percentage of sucrose synthesized is calculated based on the number of moles of C atoms contained in the leaf biomass. The demand of sucrose biosynthesis in the leaves is gradually increased from 0% to 200% to account for varying rates of Suc accumulation during fall. Each square in a grid represents an FBA simulation in which (i) the nitrogen containing constituents in 1 mol of biomass is converted to the amino acid whose one-letter abbreviation appears in the square, with the additional metabolic demand of diversion of the specified minimum amount of C to sucrose biosynthesis and storage (ii) the amino acid and the sucrose are transported through the xylem to bark; and (iii) the amino acid and sucrose are metabolized to form the 32-kDA bark storage protein. The CO_2 released in leaf or bark during this process must be released. FBA used a multicellular genome-scale model of poplar and employed BSP maximization with L_1 norm minimization as its objective function and the constraint that no nitrogen or carbon other than the pool stored in biomass is available. The optimality of the amino acids are evaluated in terms of four biologically relevant metrics, (i) the amount of BSP synthesized in the bark; (ii) the amount of photon uptake, which is zero here; (iii) the number of moles of nitrogen lost as NH_3 in the leaf and (iv) the total number of moles of nitrogen lost as NH_3 in the leaf and bark. The amino acids are arranged in a row in the decreasing order of optimality, based on each metric, within a grid for each series of simulations. The optimal amino acids and the rank of glutamine for each series of simulations are mentioned for each metric, with the position of Gln placed in a box with thick borders. The grey boxed within a grid indicates simulations that did not converge. The details and metabolic justifications of the constraints on these simulations are included in **Table 1**.

Table 5. Constraints on multi-tissue FBA for leaf N mobilization in fall

This table displays the constraints placed on the senescing leaves and the bark in the multi-tissue FBA simulations for leaf N mobilization. We permitted photon uptake without any CO_2 uptake in leaves or mixing of CO_2 pools released in bark and leaves. The constraints placed on the tissues are placed to predict the N transport efficiency of amino acids from the senescing leaves to the bark for storage as BSP.

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{\text{bark}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf N mobilization
$v^{\text{bark}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo activity	RuBisCo does not fix carbon in bark
$-1000 \leq v^{\text{bark}}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{bark}}_{NH_3} \leq 1000$	Permitted N loss as NH_3	Transported N that cannot be stored in BSP is excreted as NH_3
$0 \leq v^{\text{bark}}_{CO_2} \leq 1000$	CO_2 excretion is permitted	C in transported amino acids not assimilated in BSP excreted as CO_2
$-1000 \leq v^{\text{bark}}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
Constraints on transport	**Equations and Bounds**	**Justification**
$- \leq \leq$	N transport from leaves to bark	N is transported to the bark exclusively as amino acid, one at a time
$v^{\text{min}}_{\text{Sucrose}} \leq \leq$	Suc transport from leaves to bark	Specified minimum amount of Suc transport from leaf to bark
$v^{\text{bark to leaf}}_{CO_2} =$	CO_2 transport from bark to shoot	CO_2 generated in bark during BSP synthesis must be excreted
Constraints on leaf	**Equations and Bounds**	**Justification**
$\mathbf{S}.\mathbf{v}^{\text{leaf}} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf*N* mobilization
$v^{\text{leaf}}_{\text{RuBisCo}} = 0$	Blocked RuBisCo in leaves	RuBisCo activity drops significantly with leaf senescence
$v^{\text{leaf}}_{\text{biomass}} = -1$	1 mol of biomass uptake	Biomass is the sole source of C and N to predict parsimonious nutrient
use $-1000 \leq v^{\text{leaf}}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{\text{leaf}}_{NH_3} \leq 1000$	Permitted N loss as NH_3	N in leaf that cannot be transported to bark as amino acid is excreted
$0 \leq \leq$	CO_2 excretion only from leaf	Unclear if CO_2 is assimilated or synthesized during leaf senescence
$-1000 \leq v^{\text{leaf}}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange flux
$v^{\text{leaf}}_{h\nu} \leq$	Photon uptake is not permitted	Photons cannot be captured in leaves once RuBisCo is broken down

Table 6. Constraints on multi-tissue FBA for leaf N mobilization in fall

This table displays the constraints placed on the senescing leaves and the bark in the multi-tissue FBA simulations for leaf N mobilization. We permitted photon uptake without any CO_2 uptake in leaves or mixing of CO_2 pools released in bark and leaves. The constraints placed on the tissues are placed to predict the N transport efficiency of amino acids from the senescing leaves to the bark for storage as BSP.

Constraints on bark tissue	Equations and Bounds	Justification
$\mathbf{S}.\mathbf{v}^{bark} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf N mobilization
$v^{bark}_{RuBisCo} = 0$	Blocked RuBisCo activity	RuBisCo does not fix carbon in bark
$-1000 \leq v^{bark}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{bark}_{NH_3} \leq 1000$	Permitted N loss as NH_3	Transported N that cannot be stored in BSP is excreted as NH_3
$0 \leq v^{bark}_{CO_2} \leq 1000$	CO_2 excretion is permitted	C in transported amino acids not assimilated in BSP excreted as CO_2
$-1000 \leq v^{bark}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
Constraints on transport	**Equations and Bounds**	**Justification**
$- \leq \leq$	N transport from leaves to bark	N is transported to the bark exclusively as amino acid, one at a time
$v^{min}_{Sucrose} \leq \leq$	Suc transport from leaves to bark	Specified minimum amount of Suc transport from leaf to bark
$v^{bark\ to\ leaf}_{CO_2} =$	CO_2 transport from bark to shoot	CO_2 generated in bark during BSP synthesis must be excreted
Constraints on leaf	**Equations and Bounds**	**Justification**
$\mathbf{S}.\mathbf{v}^{leaf} = \mathbf{0}$	Mass balance without metabolite accumulation	Assumption of steady state metabolism during leaf*N* mobilization
$v^{leaf}_{RuBisCo} = 0$	Blocked RuBisCo in leaves	RuBisCo activity drops significantly with leaf senescence
$v^{leaf}_{biomass} = -1$	1 mol of biomass uptake	Biomass is the sole source of C and N to predict parsimonious nutrient
use $-1000 \leq v^{leaf}_{H,O_2,H_2O} \leq 1000$	Unrestricted exchange of H, O_2, H_2O	No reported data on H, O_2, H_2O exchange fluxes
$0 \leq v^{leaf}_{NH_3} \leq 1000$	Permitted N loss as NH_3	N in leaf that cannot be transported to bark as amino acid is excreted
$0 \leq \leq$	CO_2 excretion only from leaf	Unclear if CO_2 is assimilated or synthesized during leaf senescence
$-1000 \leq v^{leaf}_{Pi,SO_4} \leq 1000$	Unrestricted Pi and SO_4^- exchange	No reported data on Pi and SO_4^- exchange fluxes
$v^{leaf}_{h\nu} =$	Photon uptake is not permitted	Photons cannot be captured in leaves once RuBisCo is broken down

To show the optimal genome-wide pathway for N transport from leaves, we prepared a flux map of the scenario in which all the N in leaf is transported by Gln (**Figure 3.11**), with photon uptake in the leaves . The flux map predicts significant flux through the reactions carbamoyl phosphate synthetase (CPS), argininosuccinate synthase (AS), argininosuccinate lyase (AL), ornithine transcarbamylase (OTC) and glutamine synthetase (GS). Increased expression levels of these genes during leaf senescence have been reported in multiple studies (Couturier et al., 2010). In addition, it also predicts significant flux through glutamate synthase and several reactions of the tricarboxylic acid (TCA) cycle to produce 2-oxoglutarate. Recent studies have reported the significance of glutamine synthetase/glutamate synthase (GOGAT) cycle and isocitrate dehydrogenase (IDH) enzyme (Suárez et al., 2002). Thus, multi-tissue genome-scale model is fully capable of highlighting important enzymatic reactions in poplar metabolism when provided with appropriate objective function and constraints.

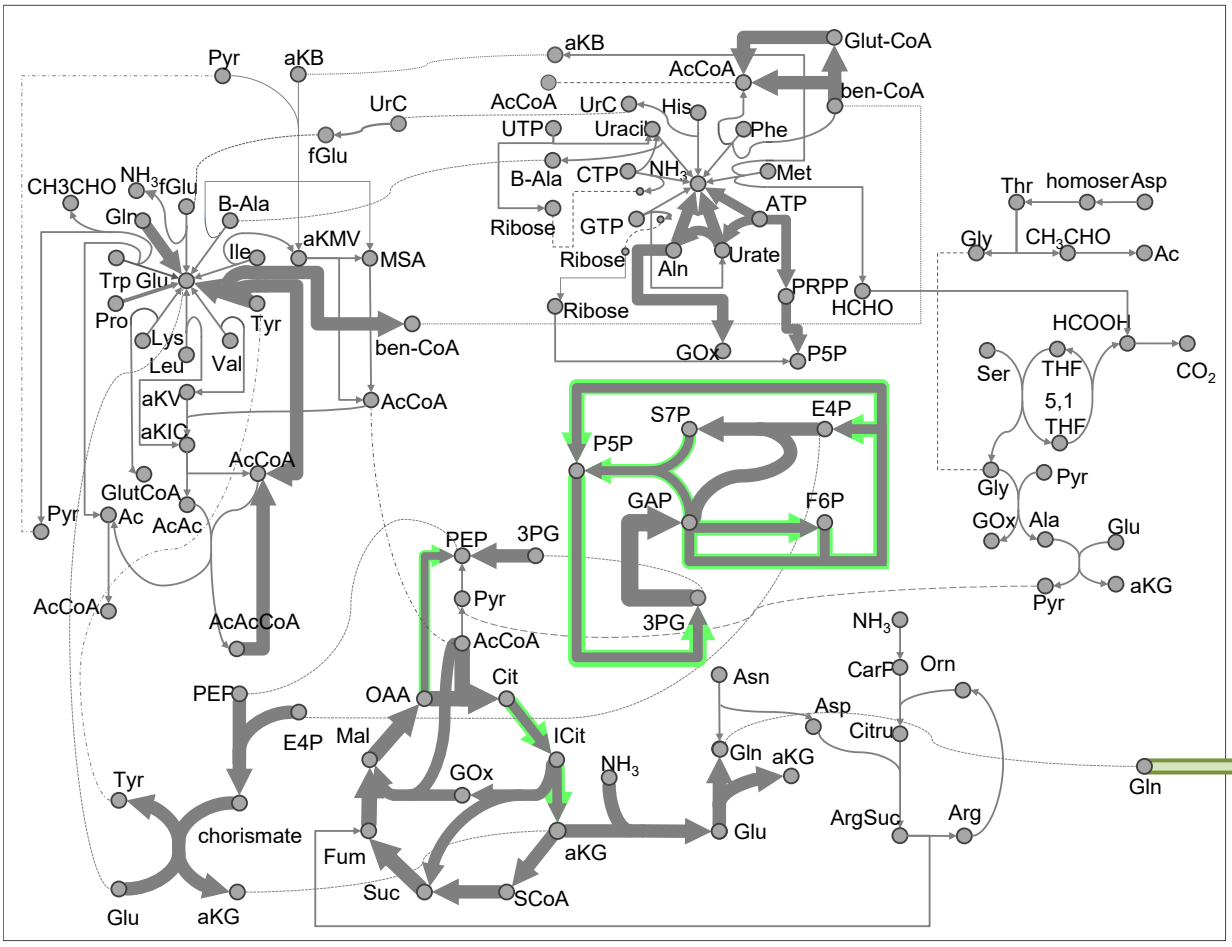


Figure 3.11. Optimal route for leaf N mobilization during leaf senescence as predicted by FBA. A, flux map displaying the reactions in the leaf cell of poplar during mobilization of leaf N content to only Arg. In an FBA simulation, the N containing metabolites in 1 mole of biomass is converted to Arg in poplar leaf; Arg is then transported to bark tissue and further metabolized there for storage as BSP. The FBA simulation was performed using the multicellular compartmented GSM and employed the objective function of maximization of BSP with L_1 norm minimization,

with the constraints of no additional N or C other than the pool stored in biomass. To aid visualization, the compartmented FBA solution was converted to a simplified uncompartmented solution, which was then used to prepare the flux map. The conversion involved three major steps, (i) removal of compartmentation from the reactions; (ii) removal of cofactors from the reactions, with the objective of observing only the C and N flow; (iii) addition of fluxes in order to calculate a total overall flux when a reaction was predicted to take place in multiple compartments. To utilize space better in the figure, multiple reactions carrying the same flux were often combined to a single overall reaction. CO_2 is treated as a cofactor for brevity, and in case of either assimilation or release of CO_2 in a reaction in a pathway, only the initial and final metabolites are shown. Each circle in the flux map represents a metabolite and each solid arrow represents a metabolic reaction or pathway. The dashed lines are used as connectors to connect a metabolite appearing multiple times in the flux map, whenever possible. The single red arrow crossing the solid boundary represents the transport of Arg to bark tissue. The arrow thicknesses represent the flux through that reaction in a logarithmic scale. The blue circle represents Glu as the nitrogen carrier in which the metabolites connected to it by solid arrows donate their N to aKG (not shown in the map), thereby forming Glu. Similarly, the two green circles represent NH_3 released in degradation pathways by the metabolites connected to them by solid arrows. The purple circle represents the donation of the N to Glu by Asn and Glu in two reactions forming Gln. A description of the schematics used to represent reactions or pathways, as used in the preparation of the flux map, is provided in panels B-G. All abbreviations of metabolites are defined in the List of Abbreviations. B, the schematic representation of a uni-uni (unireactant and uniproduct) reaction. C, the schematic representation of a uni-bi (unireactant and biproduct) reaction. D, the schematic representation of a bi-uni (bireactant and uniproduct) reaction. E, the schematic representation of a bi-bi (bireactant and biproduct) reaction. F, the schematic representation of a reactant getting converted to two products by two different reactions or pathways reaction. G, the schematic representation of two reactants getting converted to a product by two different reactions or pathways.

3.3.2 Emission of isoprene on hot days and intense sunlight

Isoprene emission by plants on hot sunny days is the major source of biogenic hydrocarbons in the atmosphere (Sharkey et al., 2008). The subsequent reaction of isoprene with hydroxyl radicals in the atmosphere leads to enhanced ozone and aerosol formation (Sharkey et al., 2008). Poplar is a major contributor to isoprene emitted by trees to the atmosphere. Isoprene is synthesized from its last precursor dimethylallyl diphosphate (DMAPP) in a reaction catalyzed by isoprene synthase (IspS). DMAPP is synthesized by two pathways, the methylerythritol phosphate (MEP) pathway and the mevalonate (MVA) pathway. In poplar, the MVA pathway is in the cytosol while the MEP pathway is in the plastid. However, the gene IspS is in the chloroplast of poplar. Isoprene emission in poplar is regulated by IspS activity and DMAPP availability in plastid (Wiberley et al., 2008).

Experimental studies carried earlier have suggested that cross talk between the two pathways is either very limited or nonexistent (Loreto et al., 2004; Rasulov et al., 2009). The contribution of the cytosolic MVA pathway to isoprene emission under most conditions is insignificant, and in the direction from plastid to cytosol, with DMAPP for

isoprene biosynthesis supplied only by the plastidic MEP pathway. The cytosolic and mitochondrial isoprenoids are mainly products of MVA pathway (Vranová et al., 2012). Additionally, the products of MVA pathway are poorly incorporated into the plastidic MEP pathway (Logan et al., 2000). Isoprene biosynthesis requires an active Calvin cycle, and is not dependent upon photorespiration (Sharkey and Yeh, 2001).
No consensus exists on why a tree emits isoprene. Suggested theories include better tolerance to heat and protection of leaves from reactive oxygen species among others (Sharkey et al., 2008). Here, we investigated with FBA (i) the theory of isoprene biosynthesis as a sink mechanism to get rid of excess energy or carbon, and (ii) if any crosstalk exists between the cytosolic MVA and plastidic MEP pathways.

We employed our single-tissue GSM to identify the preferred pathway of isoprene biosynthesis if the sole cellular objective was to send all the C flux towards isoprene. We performed the simulations with fixed CO_2 uptake, free exchange of all other substrates essential for growth and objective function of isoprene maximization with L_1 norm minimization. Additionally, we gradually increased the photon flux to the model to check if the choice of pathway is impacted by incident photon flux. FBA picked MEP pathway for isoprene biosynthesis, presumably due to lower energetic cost as well as the smaller number of enzymatic reaction steps in MEP pathway. Repeating the simulations without the L_1 norm minimization did not change the preferred pathway. We repeated the simulations with two other objective functions, maximization of NADPH and NADH production, with and without L_1 norm minimization. However, FBA did not choose MVA pathway for isoprene biosynthesis. If isoprene biosynthesis were to be a sink for excess energy captured by plants, the less energy efficient MVA pathway should have been the preferred pathway for isoprene biosynthesis. However, our FBA simulations did not indicate that.

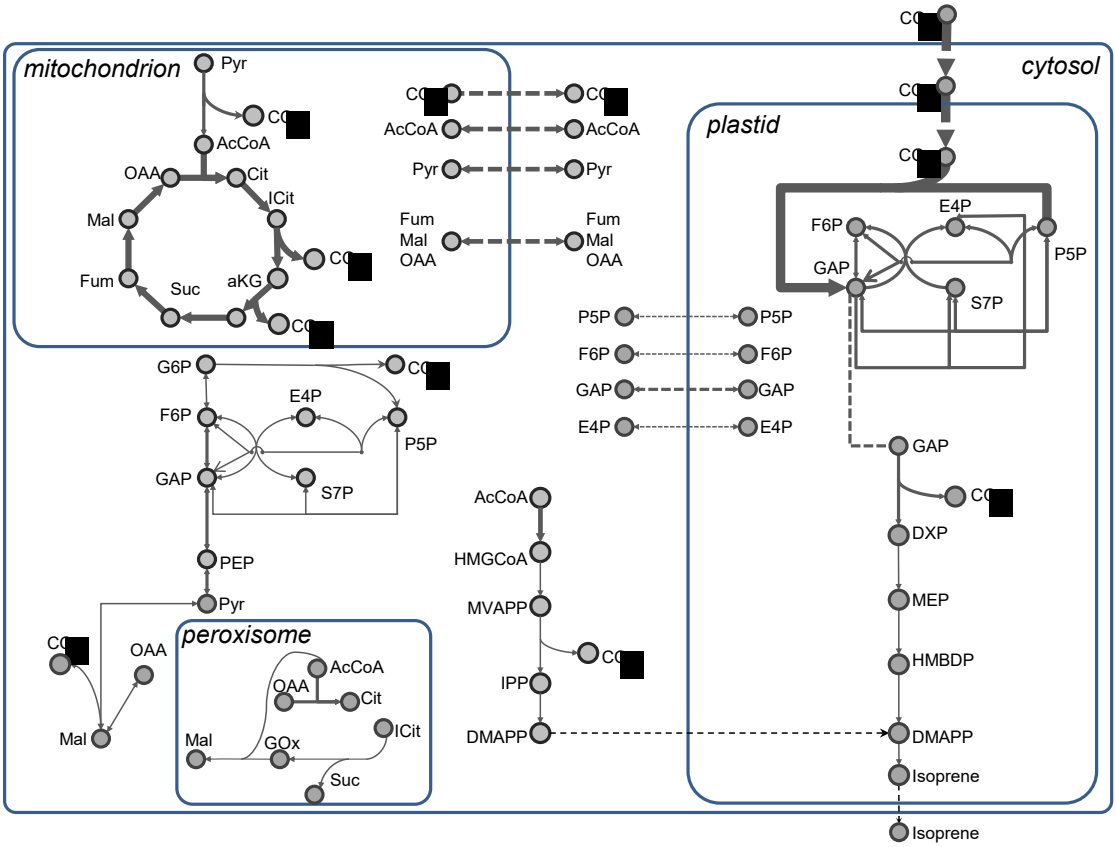


Figure 3.12. Central carbon metabolism and isoprene biosynthesis pathways of poplar. The figure displays the central carbon metabolism and isoprene biosynthesis pathways in poplar. FBA were performed with multiple objective functions, with and without L_1 norm minimization. The MEP pathway is energetically more efficient in producing isoprene than the MVA pathway, and hence it was predicted as the preferred pathway by FBA in all simulations.
We considered the simultaneous biosynthesis of biomass with isoprene, with the C flux from biomass was increasing gradually from biomass towards isoprene biosynthesis in another series of FBA. In none of the analyses MVA pathway was the preferred pathway for isoprene biosynthesis. MVA pathway carried flux only when the MEP pathway was restricted by limiting the upped bound. Based on our results we hypothesize that plants preferentially use MEP pathway to synthesize DMAPP, the precursor of isoprene, for energy energetic efficiency. Hence, unless the enzymes in MEP pathway are inhibited MVA pathway will not contribute to isoprenoid biosynthesis outside of plastid.

The reactions in central carbon metabolism (CCM) of poplar are distributed across multiple organelles (**Figure 3.12**). We investigated if blocking any central carbon metabolism intermediate exchange reaction could impact the choice of DMAPP biosynthesis pathway. However, MEP remained the predicted pathway for isoprene biosynthesis. for the biosynthesis of DMAPP in cytosol transport of glyceraldehyde-3-phosphate (GAP) to cytosol was not essential. Cytosolic isoprenoid production could be sustained by the transport of any of erythrose-4-phoshate (E4P), pentose phosphate (P5P), fructose-6-phosphate (F6P) due to the presence of cytosolic PPP. The distributed nature of central carbon metabolism provided tremendous flexibility in terms of

channeling reduced forms of carbon across organelles. However, whether they play a role in synthesis of isoprene or other isoprenoids remains unclear.

3.4 Discussion

We reconstructed a genome-scale model of a perennial woody tree and employed it to investigate biological phenomena specific to woody trees. This, to the best of our knowledge, is the first reconstruction of a tree GSM and application of predictive modeling to seek answer to a hitherto unanswered question on tree specific metabolism. Network size, dynamic nature of tree metabolism, metabolic events with a wide range of timescales and the relative lack of data to constrain a model render tree metabolism investigation with a challenging task. As a result, cellular objective of a tissue is often not apparent requiring employment of multiple different objective functions and analysis of all the predictions to check if any objective function can explain a phenomenon better than the others. These limitations necessitate incorporation of known biochemical evidence in the form of flux restrictions through specific reactions. Given the unclear objective and physiological constraints, the predictions must be analyzed in the context of biologically relevant metrics. We employed these strategies to analyze our FBA predictions. Despite these strategies, there are inherent limitations to the type of problems which can be investigated by FBA.

3.4.1 Efficiency of N mobilization for early spring growth might be a factor in selection of N transport amino acid

The selection of Gln as the predominant N transport amino acid during fall appears to be a compromise necessitated by the need for rapid early spring growth. The highly fluctuating seasonal N supply (Rennenberg et al., 2010) requires poplar to assimilate soil N while available. The optimal conditions for growth with suitable soil temperature, adequate sunlight and with available nitrogen in soil occur during late spring and summer. Thus, when N is available in soil poplar must have the photosynthetic machinery, which requires N availability, to generate the energy necessary for N assimilation. Optimal mobilization of stored N reserves to synthesize shoot biomass during early spring is critical to ensure survival of the tree.

We hypothesize that efficient N mobilization during early spring is a primary objective in the selection of N transport amino acid. Optimality in the late spring, as well as relatively better energetic efficiency during early spring might be the reason why poplar has evolved the machinery to mobilize BSP nitrogen through Gln. Leaf N mobilization as Gln, with limited photosynthesis in senescing leaves, leads to suboptimal storage of around 80% of leaf nitrogen in BSP. However, a faster growth during early spring could would be beneficial to poplar in assimilating soil nitrogen when it is available. Thus, optimal BSP mobilization could make poplar more viable even with suboptimal leaf N mobilization. However, we cannot speculate why poplar has not evolved different N transport machineries for leaf N mobilization. This question could be of potential interest for further investigation.

3.4.2 Multi-tissue GSM and careful selection of constraints is essential for capturing meaningful information on complex metabolic events

The biosynthesis and degradation pathways of many amino acids, as well as other biomass constituents, involve reaction steps that release CO_2. The fixation of CO_2 and NO_3, when available, are the major sinks of ATP and reductant cofactors. Generation of ATP and reductant cofactors is directly related to photon uptake. Thus, with the uptake of photons, the GSM can generate enough ATP and reductant to mobilize the entire leaf N content to any single amino acid. We know from experimental evidence that this is not what happens physiologically. Thus, either the uptake of photons must be limited with some arbitrary value or the impact of a constraint on the predicted solution must be analyzed systematically. Additionally, without any photon uptake Gln can metabolize around 80% of the nitrogen content in the leaf, which is supported by measurements from field studies. A single simulation or a simulation without a multi-tissue GSM fails to capture such information.

3.4.3 Interpretation of simulations based on multiple biologically relevant metrics is necessary for understanding metabolism in complex systems

We stress that the choice of judiciously placed constraints and data analysis based on multiple metrics is as important as the constraints based on physiological evidence alone. We performed simulations in which we let the model choose from all amino acids as ways for nitrogen transport from bark to shoot. None of the simulations predicted Gln as the preferred transport amino acid. Instead, FBA selected all amino acids in the same proportion as BSP as the transported amino acid profile. Thus, FBA is unlikely to provide meaningful predictions unless the problems are constructed carefully. The quality of FBA predictions could be improved by reducing the degree of freedom of the network by incorporating measured substrate uptakes and BSP formation rates. However, obtaining a large set of experimentally measured reaction flux to sufficiently constrain the metabolic network is nontrivial. Additionally, this nullifies the advantage of fast analysis and hypothesis development by FBA. Additionally, constraining a large network with a small set of experimentally measured fluxes does not guarantee a unique solution. For example, the predicted flux scenarios during leaf N mobilization are vastly different, with and without RuBisCO activity, all other constraints remaining identical. All these factors necessitated a careful construction of the FBA problems and a strategy to analyze the predictions. Without evaluation on several metrics, the analyses cannot provide meaningful and biologically relevant answers to the questions investigated. We stress that the goal in our work was not to quantify the fluxes through the entire metabolic network of poplar. We believe that FBA, a computational technique, is suitable for providing qualitative answers that cannot be answered easily with experiments.

3.4.4 Context specific objective function is essential for specialized metabolism in plant tissues

Plant cells rarely operate with only once cellular objective. Multiple cell types are involved in carrying out different tasks towards a common goal. Hence, single objective functions like maximization of biomass or maximization of ATP or NADPH cannot capture the complex metabolism in plants. Earlier plant GSMs have employed single objective functions to simulate plant metabolism. However, none of those studies

involved investigation of previously unexplained metabolic phenomena. We added pseudo reactions containing the two objective functions of interest to our model, with varying coefficients of interest (COI) between the two objectives. In our investigation, none of these objectives could explain the selection of Gln as N transport amino acid or explain how isoprene synthesis helps poplar leaves. We surmise that isoprene emission is probably not a quest for optimality in any sense by poplar leaves. The flux solution that can capture isoprene emission as a stress response possibly lies in the interior of the feasible solution space which cannot be captured accurately by linear programing. Capturing such solutions to explain transient metabolic events should be a priority in plant metabolic modeling.

3.5 Conclusion

We developed a genome-scale model of poplar and have successfully used FBA to find answers to questions on poplar metabolism. The predictions, which support field observations, have led to nontrivial hypotheses regarding questions that have remained unanswered until now. However, the predictions do not equate experimental validations. Theoretically, a plant wide nonstationary ^{13}C isotope labelled MFA can experimentally validate our predicted metabolic flux maps for N recycling. Given the costs, complexity and duration involved it is highly unlikely that an experiment of such a scale will be feasible in near future. We believe that development of computational frameworks that can incorporate competing objective functions will be better suited for tackling such complex questions.

Chapter 4 : A Genome-Scale Stoichiometric Model of *Zea mays* for Investigation Endosperm Metabolism: Role of Pyruvate Phosphate Dikinase and Amino Acid Fermentation on Energetics of Grain Filling

This chapter is part of a manuscript under preparation. Copyright maybe transferred to the publisher in the future.

Navadeep Boruah, Brendon K. Cox, Alan M. Myers and Ganesh Sriram*

NB, AMM and GS conceived this project. NB and BKC constructed and analyzed the GSM, GS provided input on analyses, AMM genetically modified the cells. NB and BKC wrote the manuscript, GS critically edited it. N.B. and B.K.C. contributed equally to this work.

Department of Chemical and Biomolecular Engineering (N. B., B. K. C., G. S.), University of Maryland, College Park, MD 20742, United States of America and Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology (A.M.M.), Iowa State University, Ames, IA 50011, United States of America

4.1 Background

Maize (*Zea mays)* is the most produced cereal grain in the world. In cereal grains, the endosperm is the nutrient storehouse of the kernel. During seed germination and early kernel development, the endosperm supports and controls embryo growth by supplying nutrients derived from the stored nutrient reserve (James et al., 2003; Méchin et al., 2007; Sabelli and Larkins, 2009; Yan et al., 2014). The predominant nutrient in the maize endosperm is starch (Emes et al., 2003; Méchin et al., 2007; Sabelli and Larkins, 2009), along with significant protein (Shukla and Cheryan, 2001; Gibbon and Larkins, 2005; Zhang et al., 2015) and a small amount of monoacyl lipids (Morrison, 1988; Shewry and Halford, 2002). Because altering the stored nutrient profile can potentially increase grain yield and quality, the maize endosperm is an attractive target for metabolic engineering (Yan et al., 2014). While the development of the maize endosperm has been extensively studied, fundamental questions on the energetics of storage product accumulation remain unresolved.

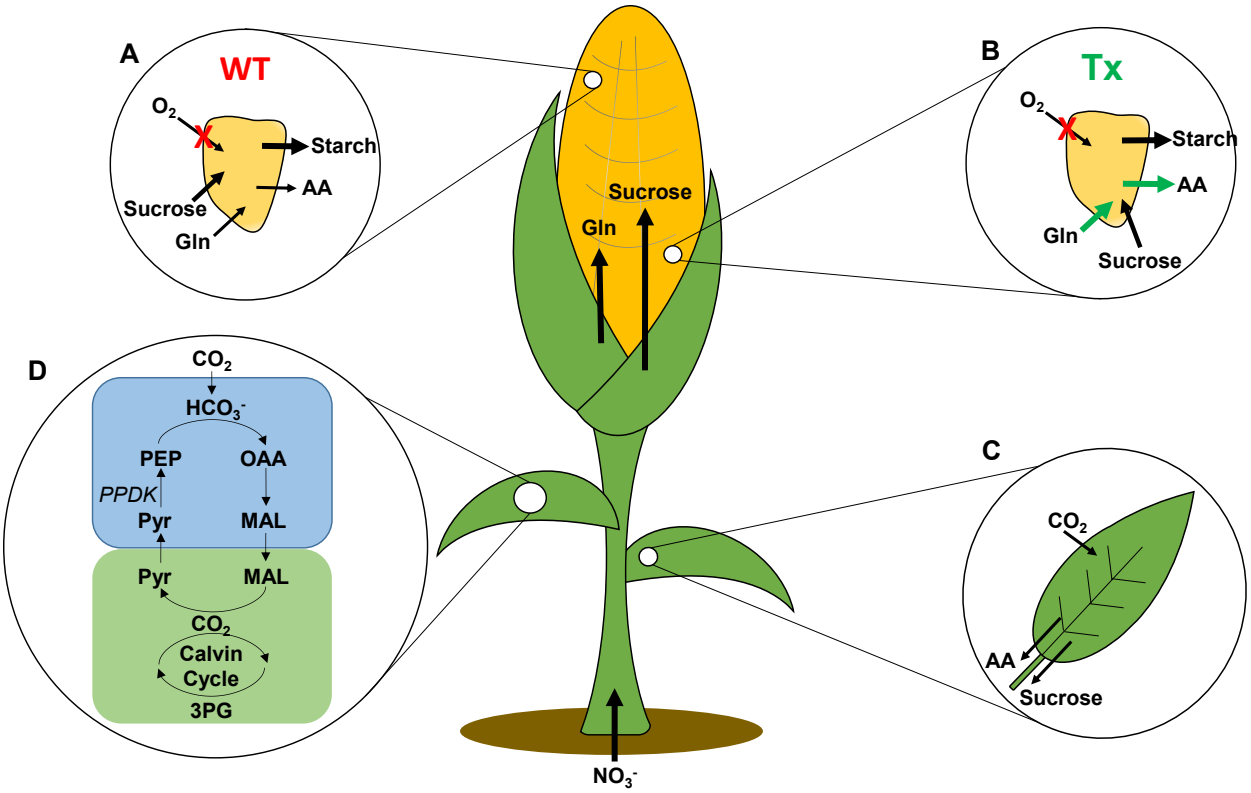


Figure 4.1. Schematic representation of nutrient flow in maize. Maize leaves photosynthetically fix CO_2 and then convert the photosynthate to sucrose. Nitrates from soil are absorbed by the roots and transported through the xylem to the leaves, prior to assimilation into amino acids. Suc and Gln are transported from the leaves to the kernel, which the endosperm metabolizes into storage products. In FBA simulations, *MaizeModel* was constrained with biologically realistic constraints specific to the hypoxic endosperm, including: (i) blocking oxygen uptake, (ii) allowing Suc and Gln as the only sources of carbon and nitrogen. The black arrows indicate the direction of metabolite flow simulated by FBA. Arrows marked with a red cross represent blocked metabolite uptake in our model. A, Nutrient uptake and storage product synthesis in WT maize endosperm with highly expressed *ppdk*. B, Nutrient uptake and storage product synthesis in ppdk deficient Tx maize endosperm. C, Suc and Gln flow from maize leave. D, Role of *PPDK* during C4 photosynthesis in maize leaves. All abbreviations of metabolites are defined in the List of Abbreviations.

In maize, endosperm development involves tightly coordinated metabolism of carbohydrates and amino acids across multiple tissues. C_4 photosynthesis in the leaves synthesizes carbohydrates from CO_2 (**Figure 4.1C**). Roots absorb nitrogen in the form of nitrates and assimilate it into amino acids (**Figure 4.1C**). Sucrose (Suc), along with amino acids, are supplied through the transport tissues. Suc (Emes et al., 2003) and Gln (Misra and Oaks, 1985) are the predominant carbohydrate and amino acid transported to the kernel (**Figure 4.1A**, **Figure 4.1B**). Amino acids other than Gln are also transported, albeit to a lesser extent (Arruda and Da Silva, 1979). Suc is hydrolyzed in the endosperm to hexose carbohydrates glucose (Glc) and Fructose (Fru). These hexoses then undergo phosphorylation to glucose-1-phosphate (Glc-1-P) and enter the starch biosynthesis pathway. Starch, which consists of α-glucan polymers amylose and amylopectin, starts

accumulating in amyloplasts around 10 days after pollination (DAP) (Sabelli and Larkins, 2009). The bulk of starch is deposited between 12 DAP to 35 DAP (Méchin et al., 2007). Similarly, the transported amino acids undergo further metabolism in endosperm to synthesize the storage proteins. Accumulation of these storage proteins, called zeins, starts around 10 DAP (Wu et al., 2010) and continues until 70 DAP (Méchin et al., 2007).

The phosphorylation of hexoses and the formation of peptide bonds in storage proteins require Gibbs energy in the form of NTPs. Hence, the endosperm is an immense energy sink during grain filling (Sabelli and Larkins, 2009). The dense endosperm is severely hypoxic, due to exclusive reliance on diffusive oxygen uptake, with a mean O_2 concentration of 1.4% of atmospheric saturation (Rolletschek et al., 2005; Borisjuk and Rolletschek, 2009). The endosperm cannot generate ATP by tricarboxylic acid (TCA) cycle as O_2 is essential for regeneration of cofactors in TCA cycle reactions (Sabelli and Larkins, 2009). Earlier studies have indicated that glycolysis plays a major role in supplying ATP in hypoxic endosperm (Méchin et al., 2007). However, glycolysis generates far less ATP per mol of Glc metabolized, compared to TCA cycle. Clearly, endosperm metabolism is flexible enough to fulfil its energetic need. The contributions of the different ATP generating reactions to endosperm ATP demand remains unknown. Additionally, the enzyme pyruvate phosphate dikinase (PPDK), which converts phosphoenolpyruvate (PEP) to pyruvate (Pyr), was observed to be highly expressed in maize endosperm and hence was suspected to play a vital role in ATP supply (Méchin et al., 2007; Chastain et al., 2011). A recent study investigated PPDK's role in ATP supply in B73 maize endosperms by silencing it with RNA interference (Lappe et al., 2018). The total quantity of starch and proteins in the PPDK silenced mutant was largely unaffected. This led to the conclusion that PPK is not essential for ATP supply in maize endosperm (Lappe et al., 2018). Thus, the sources of ATP and the associated metabolic phenomena in maize endosperm still remain unresolved and warrants further investigation.

Flux balance analysis (FBA) with genome-scale models (GSM) is the most widely used method to predict cellular metabolism. The last ten years have seen the reconstruction of multiple *in-silico* genome-scale models of leafy plants. Genome-scale reconstructions have been employed to investigate metabolism in multiple cell types of *Arabidopsis thaliana* (Poolman et al., 2009; Dal'Molin et al., 2010c; Mintz-Oron et al., 2012; Cheung et al., 2014), multiple maize (*Zea mays*) tissues (Dal'Molin et al., 2010b; Saha et al., 2011; Simons et al., 2014bSimons et al., 2014a; Seaver et al., 2015), developing rice (*Oryza sativa*) leaf cell (Poolman et al., 2013), developing barley (*Hordeum vulgare*) endosperm (Grafahrend-Belau et al., 2009), developing rapeseed (*Brassica napus*) embryo (Pilalis et al., 2011), foxtail millet (*Setaria italica*) leaf/stem (Dal'Molin et al., 2016), tomato (*Solanum lycopersicum* L.) leaf (Yuan Huili et al., 2016), sugarcane (*Saccharum officinarum*) leaf (Dal'Molin et al., 2010b) and sorghum (Sorghum bicolor) leaf (Dal'Molin et al., 2010b). The complexity of predictive analyses with GSMs have also progressed from metabolic flux predictions in cell suspension culture (Poolman et al., 2009) to predictions of resource allocation in multiple tissues (leaf, stem and root) in Arabidopsis (Dal'Molin et al., 2015). An excellent review on the state of plant genome-scale modeling has been published recently (Dal'Molin and Nielsen, 2012).

Multiple groups have employed maize GSMs to investigate a wide range of metabolic scenarios. These include (i) photosynthesis and flux partitioning between mesophyll and bundle sheath cells (Dal'Molin et al., 2010b; Saha et al., 2011; Simons et al., 2014bSimons et al., 2014a); (ii) characterization of metabolic capacity (Saha et al., 2011) and (ii) assessment of nitrogen (N) use efficiency (Simons et al., 2014bSimons et al., 2014a). The sizes of the models have increased progressively with each update. The first maize GSM, C4GEM contained 1588 unique metabolic reactions. The latest maize GSM reconstructed by Simons et al. from the MetaCrop and MaizeCyc databases, prior to our reconstruction contained 8525 non-unique reactions distributed between two cell types. Different metabolic pathway databases are not always consistent, often differing in enzyme annotations which is dependent upon the computational pipeline employed in enzymatic function assignment and manual curation (Walsh et al., 2016). Thus, utilization of multiple databases while creating a GSM can provide a model with greater coverage and higher accuracy (Walsh et al., 2016). Since the publication of the last maize GSM, the MaizeCyc (Monaco et al., 2013) and CornCyc (http://www.plantcyc.org) databases have undergone revisions. Hence there is a need for an updated maize GSM.

Here, we reconstructed a fully-compartmentalized maize GSM from the CornCyc 7 (http://www.plantcyc.org) and MaizeCyc 2.2 (Monaco et al., 2013) databases. These two databases are based on the same maize genome assembly version (B73 RefGen_v2), but enzymatic function assignments were performed by different groups (Walsh et al., 2016). The combination of the two databases has resulted in a plant GSM with unprecedented depth, with 4448 unique metabolic reactions and 3824 unique metabolites for a single cell. This fully curated model can produce a wide range of major biomass constituents individually as well as together in a wide range of proportions. The GSM can be readily employed to make predictions with the COnstaints Based Reconstruction and Analysis (COBRA) toolbox in MATLAB (Schellenberger et al., 2011). Additionally, tissue specificity can be rendered easily by appropriate selection of a subset of the reactions in our model (Opdam et al., 2017). We employed FBA and flux variability analysis (FVA) with this non tissue-specific GSM to investigate endosperm metabolism in two maize lines, wild type B73 maize and a *ppdk* deficient mutant, during starch filling. Analyses of the experimental data and simulation results point towards a dependence of starch filling on amino acid biosynthesis, the concurrence of which is well known (Sabelli and Larkins, 2009). Additionally, we explain some of the spatial variations in metabolite profiles in maize endosperms (Gayral et al., 2017) in the light of our simulations.

4.2 Materials and Methods

4.2.1 Genetic modification of maize and kernel collection

The Tx plants containing one copy of PPDK-RNAi transgene were crossed to wild type B73 plants containing no transgene. The progenies on these ears from this cross segregated 1:1 for Tx to WT. The Tx progeny lack any PPDK protein or enzymatic activity, with the WT progeny the B73 wild type siblings. All individuals develop on the same ear so there is no plant to plant variation.

The maize ears from field grown plants were harvested at 20 DAP and immersed in liquid nitrogen to quench all metabolism. While frozen, the kernels were removed

quickly from the ear and the endosperm was stored at -80 ℃ following the removal of the pericarp and embryo the embryo were removed. The genotypes were determined later by RNA extraction followed by PCR to identify which endosperms are Tx/+ and which are +/+. The frozen endosperms from -80 ℃ freezer were lyophilized in dry ice/ethanol bath for 3 days.

4.2.2 Metabolite extraction and analysis

The lyophilized endosperms were sliced and a small sliced section from each endosperm was collected for metabolite extraction. This slice was collected in a 2 mL centrifuge tube and was powdered by using a vortex mixer. We added 0.7 mL boiling methanol and 25 µL of deionized H_2O to each sample and incubated at 30 ℃ for 30 mins. Following incubation, we added 0.7 mL of deionized H_2O and 0.37 µL of chloroform to each sample. The samples were vortexed and centrifuged at 13,300 min^{-1} at 4 ℃ for 10 min. We collected the upper methanol/H_2O layers containing the polar soluble metabolites and evaporated them under a stream of nitrogen. We discarded the lipid containing chloroform layers and dried the remaining pellets under a stream of N_2.

We dissolved the soluble metabolite samples in 200 µL of deionized H_2O, transferred to a 300 µL of GC vial and lyophilized overnight. Then we reconstituted the lyophilized residues in 125 µL of N,N-dimethylformamide (DMF), and incubated at 30 ℃ for 90 min after addition of 10 µL of 20 mg/ML methoxyamine hydrochloride in pyridine solution. We drew volumes of 50 µL from each of these solutions to two separate 300 µL GC vials for derivatization. We employed two derivatization methods as not all metabolites can be measured accurately by just one. The contents of one vial was derivatized with 50 µL of MTBSTFA + 1% TBDMCS MTBSTFA +1% TBDMS (N-methyl-N-(tert-butyldimethylsilyl)trifluoroacetamide, tertbutyldimetheylchlorosilane) (Sigma-Aldrich, St. Louis, MO) at 70 ℃ for 90 min, and the other was derivatized with 50 µL of TMS (N,O-Bis(trimethylsilyl)trifluoroacetamide) (Sigma-Aldrich, St. Louis, MO) at 30 ℃ for 90 min. These derivatized samples were injected into a Varian 300 GC-MS.

We hydrolyzed the dried pellets remaining from the previous step in 3 mL of 6N HCl and incubating at 155 ℃ for 12 hours. Post acid addition, we flushed the hydrolysis tubes with N_2 gas to remove any residual O_2 and evacuated the tubes by using a vacuum pump. This procedure was repeated five times. We evaporated the hydrolysate in a stream of N_2 and reconstituted the dried samples in 125 µL of deionized H_2O. We drew volumes of 50 µL from each of these solutions to two separate 300 µL GC vials, which were lyophilized, and reconstituted in 125 µL of DMF, and then derivatized with 50 µL of TMS and 50 µL of TBDMS as described earlier.

4.2.3 Construction of the draft GSM

The data files obtained from the *CornCyc* and *MaizeCyc* databases were mined with in house MATLAB scripts to obtain two datasets containing the entire inventory of reactions and metabolites in the two databases. The two databases had some overlap with different numbers of reactions and metabolites. The extracted data sets were combined to reconstruct a single model, after ensuring that the overlapping reactions and metabolites appear only once. However, this model was FBA-ready and need extensive curation. We curated the model by utilizing the strategies described in the paragraphs below. The

subcellular compartmentation of reactions was assigned by scouring existing literature, with a major part of the data mined from Simons et al. (Simons et al., 2014bSimons et al., 2014a). A failure to assign a compartment to a reaction resulted in the reaction being assigned to cytosol. Spontaneous reactions that did not require an enzyme were assigned to all compartments. The biomass synthesis reaction was assigned to the cytosol and the extracellular transport reactions were added to the model. **Figure 4.2** depicts the overall workflow for model reconstruction and refinement. A good review of methods employed in construction and curation of GSM can be found in our research group's previous publications (Simons et al., 2014a).

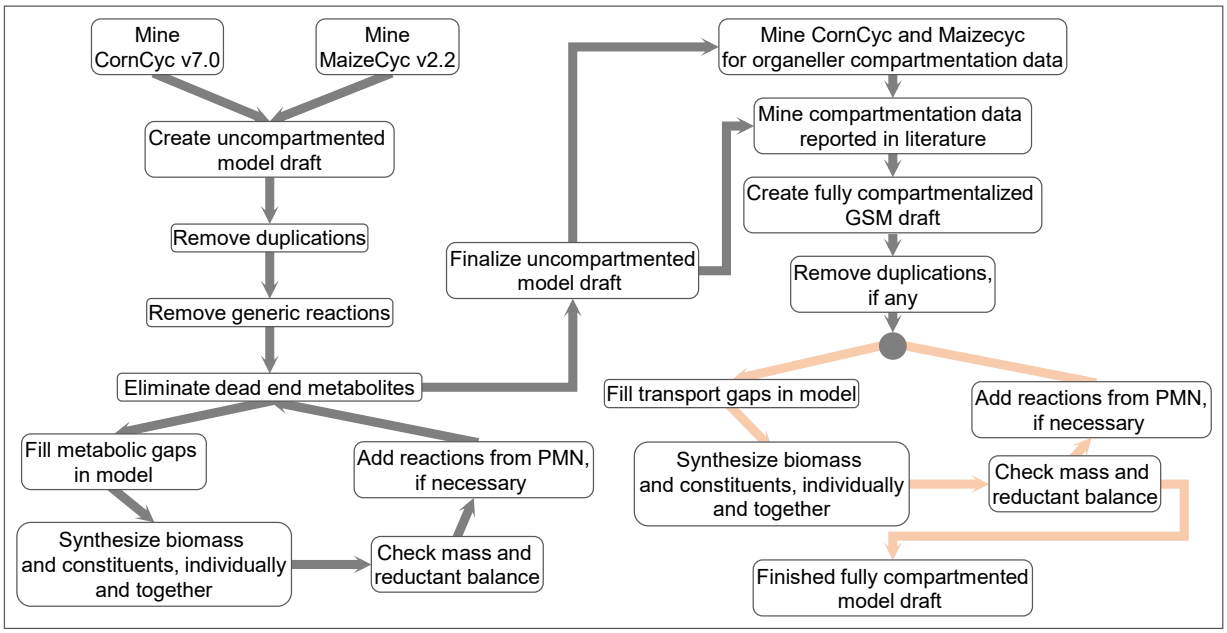


Figure 4.2. Workflow of *Zea Mays* genome-scale model reconstruction. We reconstructed a draft of *Zea mays* GSM *MaizeModel* by combining reaction information from the *MaizeCyc* and *CornCyc* databases. We then curated this uncompartmented draft model extensively with MATLAB scripts developed in-house. Curation involved removal of duplicate of reactions and metabolites as well as expansion of generic reactions into species-specific versions. To identify metabolic gaps and to remove redundant ATP- and reductant-generating substrate cycles in the model, we utilized an iterative model curation method employing FBA. We incorporated subcellular compartmentation and inter-organelle transporters into the model by mining the two databases and other published literature extensively. We identified transport gaps and energy-generating substrate cycles, introduced due to compartmentation, in the model and eliminated them iteratively employing FBA. The final draft of this model contains more unique metabolic reactions than any maize GSM published thus far. In this manuscript, we report the analysis of this model to investigate the metabolic landscape in maize endosperm during grain filling. The reconstructed specific fully curated GSM can produce all of the major biomass constituents individually as well as together in a range of proportions. The GSM can be readily employed to make FBA predictions with the COnstaints Based Reconstruction and Analysis (COBRA) toolbox in MATLAB (Becker et al., 2007; Schellenberger et al., 2011). Finally, we did not attempt to render tissue specificity to our model. We believe

that tissue specificity can be readily achieved by appropriate selection of a subset of all the reactions/pathways using an omics integration method of choice by the user. Additionally, the investigation of other metabolic phenomena in maize might require further curation of the model in the form of addition of metabolic reactions and inter-organelle transport reactions.

4.2.4 Gap filling and substrate cycles

The gaps in the GSM were identified by performing the FBA simulations for production of biomass. Close inspection of the model following failed simulations revealed the gaps. The gaps were fixed multiple strategies in the following order of preference: (i) addition of reactions with experimental evidence of existence in C4 plants, specifically cereals; (ii) addition of reactions from any known plant pathway in *MetaCyc*, as long as all the reactants and products in the reaction are already present in the model; (iii) addition of reactions that had been observed experimentally in cereal endosperm during hypoxic periods. Reactions producing carbohydrate biomass constituents were added and mass balanced in order to artificially produce carbohydrates such as galactose and xylose for the biomass equation. This was done to ensure that the FBA analyses can capture a realistic scenario with possible interventions.

Majority of transport reactions between compartments were mined from Simons et. al. (Simons et al., 2014a) and other published literature. We did not attempt to comprehensively include all extracellular transport reactions, but we allowed excretion of all biomass constituents. Thus, the model contains only those extracellular transport reactions essential to perform the FBA analyses in this paper.

The substrate cycles in the model led to mass and energy imbalances leading to unrealistic production of biomass and its precursors. Majority of substrate cycles contributing to mass imbalances were removed by correct elemental mass balancing of all the reactions in the model. The substrate cycles producing energy imbalances, by generating excess ATP and reductant, were removed by constraining directionalities of the reactions involved. Substrate cycles that were mass and energy balanced but were not experimentally observed and involved the cycling of unlikely metabolite were also removed. To the best of our knowledge, the GSM is completely free of substrate cycles capable.

4.2.5 Generic Reactions and Dead-End Metabolites

Generic reactions with R-groups in the metabolites were replaced with reactions involving specific metabolites. Often a representative metabolite was used instead of multiple specific metabolites. Reactions that used a generic electron donor and acceptor were removed or replaced with their proper cofactor. These reactions were removed in the absence of any specific metabolite. The pathways involving polymerization of a monomer, for example starch biosynthesis, were replaced with a single reaction with the monomer as the sole reactant and the polymer as the sole product. The dead-end metabolites were identified by analyzing the stoichiometric matrix of the model and were removed from the model.

4.2.6 Development of the biomass reaction

The biomass equation was obtained from Saha et. al. (Saha et al., 2011) and tailored to fit our model. Inorganic metabolites such as minerals and vitamins were not included as constituents. For fatty acid production, only palmitate and stearate were considered. All other biomass constituents from the Saha equation were included in our biomass equation. ATP cost of producing biomass was not included with our biomass reaction, so that ATP cost was only defined by the production of constituents in the network. To determine that our biomass equation was mass and energy balanced, we analyzed each biomass constituent and fixed any imbalance that was found. However, our biomass equation is specific to maize leaves, so adjustment of the biomass equation will be necessary when analyzing other tissues in maize.

4.2.7 Objective functions and analyses

FBA uses linear programming to solve the following problem:

$$\begin{aligned} \max \quad & c_j v_j \\ \text{s.t.} \quad & S_{ij} v_j = 0 \\ & v_{j,min} \le v_j \le v_{j,max} \end{aligned}$$

This is the primal of the linear programming problem, with maximization of biomass as the primary objective and L_1 norm minimization as the secondary objective. The network is artificially constrained by setting boundaries on the upper and lower bounds of reactions to reflect environmental and physiological conditions. In order to force nutrients into the system, negative values are placed on the lower bounds.

FVA uses linear programming to solve the following problem:

$$\begin{aligned} \max/\min \quad & v_{j,\max} / v_{j,\min} \\ \text{s.t.} \quad & S_{ij} v_j = 0 \\ & v_{j,min} \le v_j \le v_{j,max} \end{aligned}$$

The primary objective is always maximization of biomass, but for this problem, the secondary objective is maximizing and minimizing a single reaction while still achieving the optimal solution. The network is artificially constrained in the same way as FBA.

4.2.7 Software

All the simulations with the GSM were performed by using the COBRA toolbox in a MATLAB environment. The open source Gnu Linear Programming Kit was the solver used in the simulations.

4.3 Results

4.3.1 Knockdown of *ppdk* changes endosperm metabolite profiles

We measured cell pellet hydrolysates of endosperms of B73 maize (WT), and a mutant of B73 deficient in PPDK through RNAi (Tx), and analyzed them with gas chromatography

mass spectrometry (GC-MS). The TICs of proteinogenic amino acids, normalized to per unit dry mass of endosperm, show significantly higher amounts (**Figure 4.3**) of Gly (P-value = 0.041), Val (P-value =0.040) and Pro (P-value =0.001). No significant differences in amounts of levulinic acid (Lva), the product resulting from acid hydrolysis of hexose carbohydrates (P-value = 0.267) and starch, and Gln+Glu (P-value = 0.786) were detected in the two endosperms. The nearly identical amounts of starch synthesized in WT and Tx endosperms is also in agreement with other experimental study (Lappe et al., 2018).

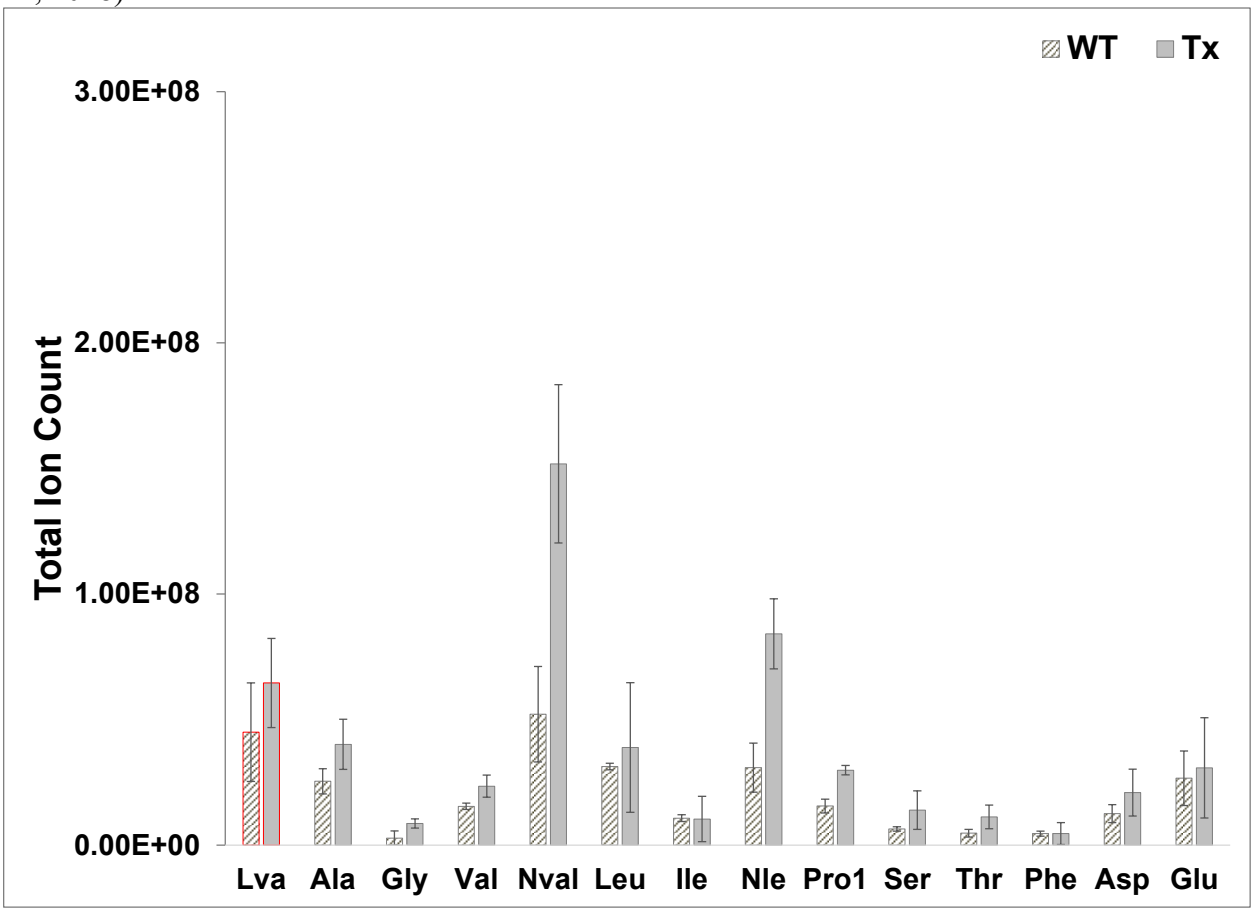


Figure 4.3. Total ion counts of metabolites detected in WT and Tx endosperms. We hydrolyzed the pellets from WT and TX endosperms, after extraction of soluble metabolites, in acid. The samples then were derivatized and injected to a GC-MS. We identified the metabolites and quantified the total ion counts (TIC) under the metabolite peaks with Varian MS Workstation software. The comparison of the TICs of the metabolites was based on per mg dry weight of the endosperm samples. The metabolites were quantified by using TBDMS and TMS (only LVA, in red) as the derivatizing agents. The samples were injected in to a Varian 300 MS coupled to a Bruker 450 GC at a split ratio of 1:15 of injected sample to carrier gas. All abbreviations of metabolites are defined in the List of Abbreviations.

4.3.2 FBA of PPDK deficient Tx endosperm predicted altered Gln uptake and amino acid biosynthesis

We employed *MaizeModel* with FBA to assess the role of PPDK in starch filling. We allowed Suc and Gln uptake by *MaizeModel* to account for the major carbohydrate and amino acid sources of endosperm respectively (Sodek and da Silva, 1977; Emes et al., 2003). We fixed the uptake of Suc at 1 mol to assess efficiencies of starch deposition and ATP synthesis associated with it. Due to the absence of experimental data of uptake rates of Gln and other inorganic nutrients, we kept them unrestricted. In the light of experimental evidence of hypoxia, we blocked O_2 uptake (Rolletschek et al., 2005; Borisjuk and Rolletschek, 2009). To capture the enzymatic parsimony during starch filling, we employed starch maximization with L_1-norm minimization as the objective. Additionally, we permitted excretion of all major biomass constituents (amino acids, nucleotides, carbohydrates and specific organic acids) to identify potential excretory products. These constraints were employed to simulations of both WT and Tx endosperms. The additional constraints of cytosolic and plastidic PPDK reactions blockage simulated its knockdown in Tx endosperms.

We performed FBA with the constraints described earlier and inspected the flux solutions to eliminate biologically unrealistic predictions. The inspections revealed FBA predicted fluxes through the glyoxylate cycle reactions of malate synthase and isocitrate lyase. We blocked these reactions due to toxicity of glyoxylate cycle (Dellero et al., 2016) and lack of any prior experimental evidence of its activity in maize endosperm. Additionally, FBA predicted ethanol (EtOH) and glycerol-3-phosphate (Gly3P) excretions. We blocked EtOH excretion due to toxicity and Gly3P excretion due to creation of a phosphate sink. These additional constraints were imposed on all subsequent simulations.

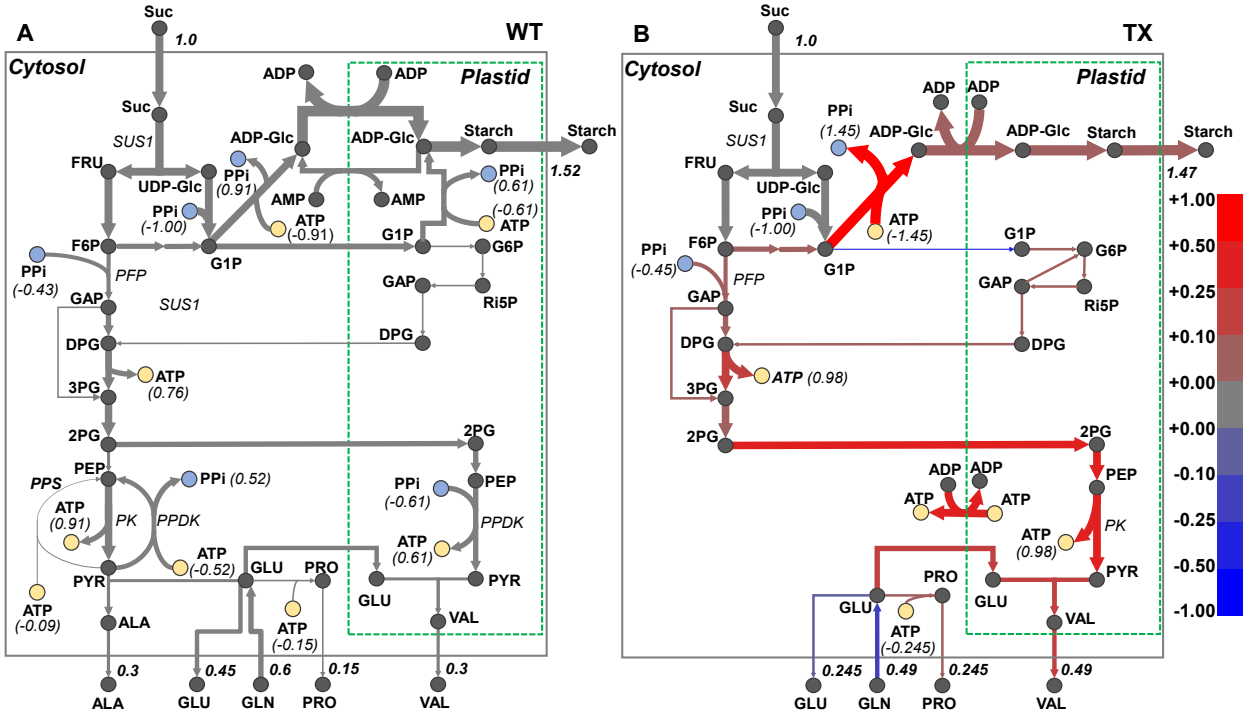


Figure 4.4. Flux balance analysis of starch filling in WT and Tx maize endosperm. We investigated starch filling in WT and Tx maize endosperms with FBA. Each simulation had constraints of 1 mole of Suc uptake, unlimited Gln uptake and blocked O_2 uptake. FBA predicted that the endosperms metabolized the nutrients to synthesize starch, along with some amino acids. Glycolysis generated the ATP needed for starch filling. Biosynthesis of amino acids (Ala, Glu, Pro and Val) expended Pyr and reductant cofactors generated by glycolysis, thereby permitting continual generation of ATP needed for starch synthesis. In WT endosperm, ADP-Glc synthesis was distributed between cytosol and plastid. The loss of *PPDK* in Tx endosperms resulted in ADP-Glc synthesis exclusively in the cytosol. Additionally, FBA predicted increased synthesis of Pro and Val in Tx endosperm, along with a small (3%) decrease in starch synthesis. Solid arrows represent the reactions and their directions. Arrow thickness encapsulates reaction flux in a logarithmic scale. Numbers next to the arrows show the reaction fluxes. Yellow circles represent ATP, light blue circles represent PPi and gray circles represent all other metabolites, with the metabolite names next to the circles. Numbers within parentheses, next to the circles indicating ATP and PPi, show the amounts synthesized or consumed in the reactions. Colored arrows indicate reaction fluxes in the Tx endosperm relative to the flux of that reaction in the WT endosperm, with the colors indicating the extent of change as per the colorbar. For the sake of brevity, the figure does not show all reactions and metabolites. A, metabolic flux in WT maize endosperm during starch filling. B, metabolic flux in Tx maize endosperm during starch filling. All abbreviations of metabolites are defined in the List of Abbreviations.

We repeated the FBA simulations of WT and Tx endosperms with the aforementioned constraints and observed inactivity of TCA cycle in the flux solutions. Predicted starch synthesis in both maize lines were nearly identical (**Figure 4.4**). Additionally, Gln uptake and amino acid excretions (Ala, Glu, Val and Pro) were predicted in both endosperms. In Tx endosperm, Gln uptake decreased (-18%) while biosynthesis of Val (+39%) and Pro (+39%) increased, compared to WT endosperm (**Figure 4.4**).

4.3.3 Gln uptake and specific amino acid biosynthesis crucial to maximal starch synthesis under hypoxia

We performed flux variability analysis (FVA) (Mahadevan and Schilling, 2003) with *MaizeModel* to investigate the rigidity of amino acid metabolism, observed in FBA simulations, and to assess if it is of any importance during starch filling. We define rigidity as invariability of a reaction flux, between maximization and minimization of flux of that specific reaction, in FVA. In FVA of the endosperms, we imposed the respective FBA predicted optimal starch excretion fluxes as additional constraints on *MaizeModel*, and then maximized/minimized the fluxes through all of the reactions. We analyzed the FVA results to identify rigidity of (i) nutrient uptake reactions, (ii) reactions in the central carbon metabolism and (iii) biomass constituent excretion reactions.

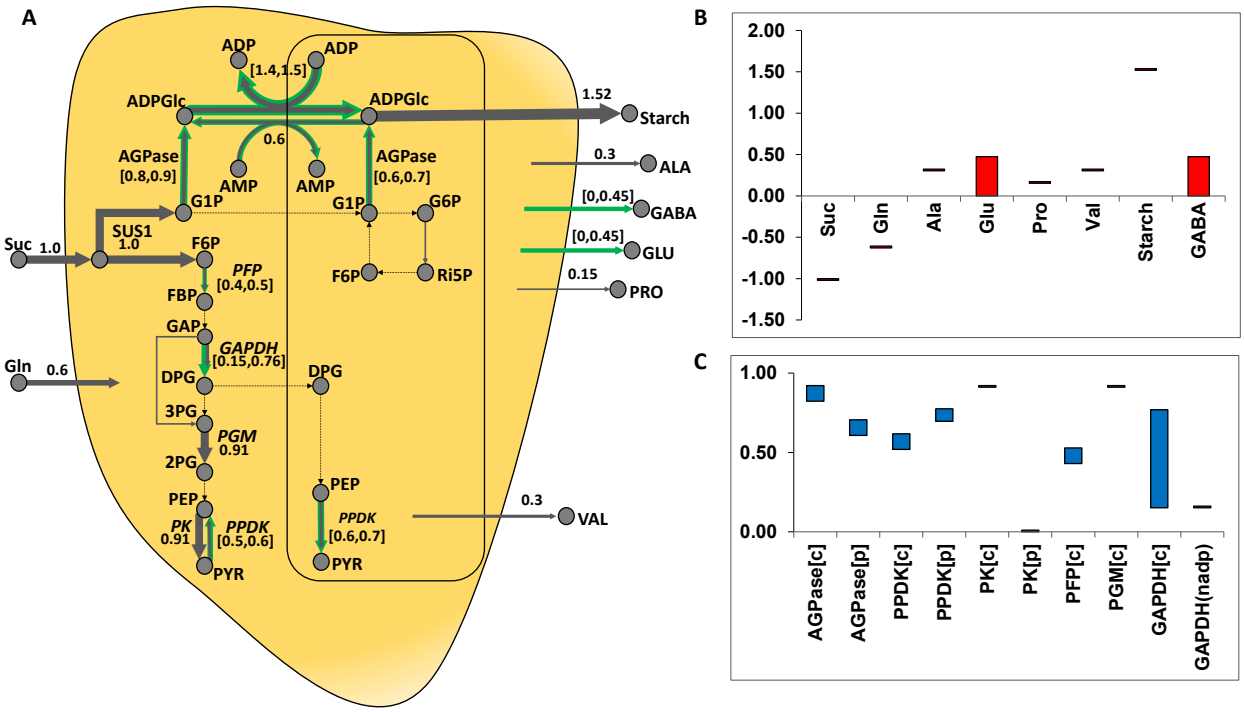


Figure 4.5. Flux variability analysis of starch filling in WT maize endosperm. We used FVA to identify metabolic rigidity during optimal starch filling in WT endosperm. We constrained all simulations with 1 mol of Suc uptake, unlimited Gln uptake, blocked O_2 uptake and the additional constraint of optimal starch synthesis. FVA predicted that for optimal starch synthesis in WT endosperm, the following fluxes are rigid: (i) uptake flux of Gln; (ii) excretion fluxes of Ala, Pro, Val; and (iii) enzymatic reactions of cytosolic PK, PGM and GAPDH(NADPH). Excretion of Glu and GABA were nonessential. A, flux variability analysis of WT endosperm metabolism. Circles represent metabolites, while solid arrows represent reactions and their directions. In case of overlapping arrows, the green arrows depict the maximum flux while grey arrows depict the minimum flux of the reaction. Gray arrows represent reactions without any flux variability. For the sake of brevity, the figure does not show all reactions and metabolites. The dotted arrows portray reactions that did not display any change in flux values. Arrow thickness encapsulates reaction flux in a logarithmic scale. The numbers within brackets, next to the arrow, display the minimum and maximum possible flux through that reaction from the FVA simulations, respectively. A single number next to a reaction display the flux of a rigid reaction. For ease of display, the dotted arrows depict multiple reactions without any flux variability as a single reaction. B, the minimum and maximum possible values of nutrient uptake and metabolite excretion fluxes. The X axis contains the flux value and the Y axis contains the metabolites. C, the minimum and maximum possible values of fluxes through important reactions. The X axis contains the flux value and the Y axis contains the enzyme names. All abbreviations of metabolites are defined in the List of Abbreviations.

FVA of WT endosperms predicted that only a small number of metabolites (Ala, Glu, Pro, Val, and GABA) could be excreted out of the endosperm (**Figure 4.5B**). Out of these, excretions of Ala, Pro and Val (**Figure 4.5B**) were rigid, along with Gln uptake. Additionally, Glu and GABA had a minimum excretion flux of zero (**Figure 4.5B**). Flux

through SUS1 and AGPase, key enzymes in starch biosynthesis pathway, were rigid and nearly rigid respectively. The majority of enzymes in central carbon metabolism were not rigid with the exception of diphosphate—fructose-6-phosphate 1-phosphotransferase (PFP), phosphoglucomutase (PGM), PPDK and pyruvate kinase (PK).

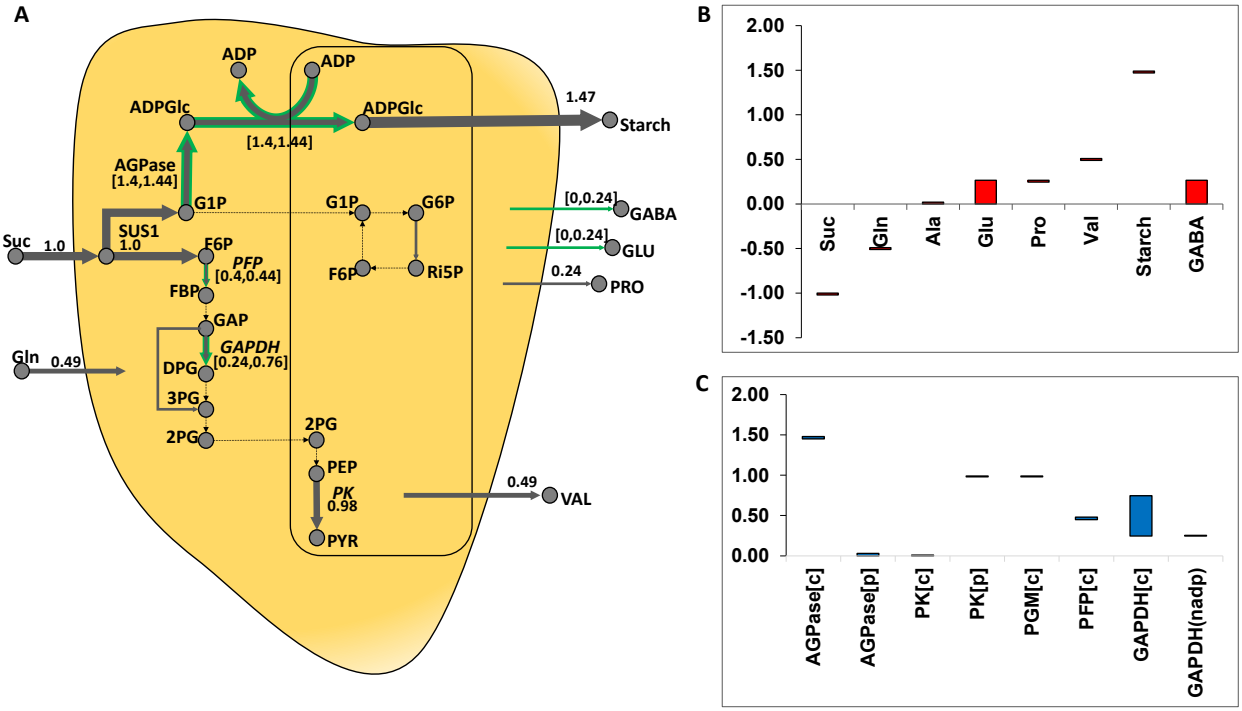


Figure 4.6. Flux variability analysis of starch filling in Tx maize endosperm. We used FVA to identify metabolic rigidity during optimal starch filling in Tx endosperm. We constrained all simulations with 1 mol of Suc uptake, unlimited Gln uptake, blocked O_2 uptake and the additional constraint of optimal starch synthesis. FVA predicted that for optimal starch synthesis in Tx endosperm, the following fluxes are rigid: (i) uptake flux of Gln; (ii) excretion fluxes of Ala, Pro, Val; and (iii) enzymatic reactions of PK (plastidic and cytosolic), cytosolic PGM and cytosolic GAPDH (nadp). Additionally, AGPase (plastidic and cytosolic) and cytosolic PFP were nearly rigid with a very limited range of possible flux values. Excretion of Glu and GABA were nonessential. The loss of PPDK reduced the redundancies in the network and limited ADP-Glc synthesis nearly exclusively to cytosol. A, flux variability analysis of Tx endosperm metabolism. Circles represent metabolites, while solid arrows represent reactions and their directions. In case of overlapping arrows, the green arrows depict the maximum flux while grey arrows depict the minimum flux of the reaction. Gray arrows represent reactions without any flux variability. For the sake of brevity, the figure does not show all reactions and metabolites. The dotted arrows portray reactions that did not display any change in flux values. Arrow thickness encapsulates reaction flux in a logarithmic scale. The numbers within brackets, next to the arrow, display the minimum and maximum possible flux through that reaction from the FVA simulations, respectively. A single number next to a reaction display the flux of a rigid reaction. For ease of display, the dotted arrows depict multiple reactions without any flux variability as a single reaction. B, the minimum and maximum possible values of nutrient uptake and metabolite excretion fluxes. The X axis contains the flux value and the Y axis contains the metabolites. C, the minimum and maximum possible values of fluxes through important reactions. The X axis contains the

flux value and the Y axis contains the enzyme names. All abbreviations of metabolites are defined in the List of Abbreviations.
FVA of Tx endosperm predicted that Gln uptake was rigid for optimal starch synthesis. Pro and Val excretions stayed rigid. Out of the amino acids that could be excreted, Glu and GABA did not require a nonzero flux (**Figure 4.6**). Additionally, cytosolic and plastidic PK stayed rigid (**Figure 4.6**). Plastidic oxidative pentose phosphate pathway (ox-PPP) and cytosolic NADP-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH(nadp)) fluxes were rigid in both endosperms (**Figure 4.5**, **Figure 4.6**). Amino acid biosynthesis used the NADPH generated using these enzymes. A loss of pgd3 gene activity, which represents expression of plastidic 6-phosphogluconate dehydrogenase (6PGDH), has been reported to decrease starch accumulation in maize (Spielbauer et al., 2013).

4.3.4 PPDK was predicted to provide flexibility of ADP-Glc synthesis across compartments

Our FBA of WT endosperm predicted that ADP-Glc synthesis by ADP-Glc pyrophosphatase (AGPase) was distributed between cytosol and plastid (**Figure 4.4**). A counter-exchange shuttle involving ADP and AMP for starch formation (**Figure 4.4**) transported ADP-Glc across the plastid membrane. Synthesis of ADG-Glc in multiple compartments (Huang et al., 2014) and transport of ADP-Glc, involving shuttle with ADP/AMP, across the plastidic membrane has been previously reported (Bowsher et al., 2007). Absence of PPDK, in Tx endosperm, led to FBA prediction of ADP-Glc synthesis exclusively in cytosol. FVA confirmed that in Tx endosperm ADP-Glc cannot be synthesized in the plastid optimally without any decrease in starch synthesis. Additionally, FBA predicted that elimination of plastidic ADP-Glc synthesis led to inactive PPDK in cytosol and plastid.
The simulations predicted PPDK to be nonessential for starch filling. Predicted starch synthesis in Tx endosperm was only marginally less (-3.2%) than in WT endosperm. This agrees with previous reports (Lappe et al., 2018). In WT endosperm, FBA predicted PPDK to (i) to consume ATP in the cytosol and (ii) to generate ATP in the plastid. Simulations also predicted PPDK as essential for recycling of PPi, with AMP, to ATP. The maize genome does not contain a cytosolic pyrophosphatase and there no evidence of a PPi transporter across the plastid membrane. Additionally, PFP, which recycles PPi, is located only in the cytosol. Hence, plastidic AGPase generated PPi cannot be recycled in the absence of PPDK. In the absence of PPDK, PK and PGM were the ATP generating reactions in endosperm (**Figure 4.4B**).

4.3.5 Pro biosynthesis is predicted to be essential for Gln uptake reduction

Zeins, the storage proteins of maize, contain unusually high amounts of Pro (Wang et al., 2014). Pro is suspected to play a regulatory role in cell cycle transition. Repression of Pro biosynthesis results in decreased starch and protein accumulation in maize endosperm (Wang et al., 2014). To identify if Pro excretion confers any metabolic advantage, we blocked Pro excretion and performed FBA with maximization of starch synthesis. Here, FBA predicted a small decrease (6%) in starch synthesis and a large increase in Gln uptake (**Figure 4.7E**). Biologically, this implies a higher amount of Gln transport to the maize kernel, and likely higher nitrogen demand by a maize plant overall. FVA with these constraints confirmed that this increased Gln uptake is rigid and is necessary to

achieve the optimal solution. Pro's potential role in minimizing N demand by maize kernel could be of potential interest.

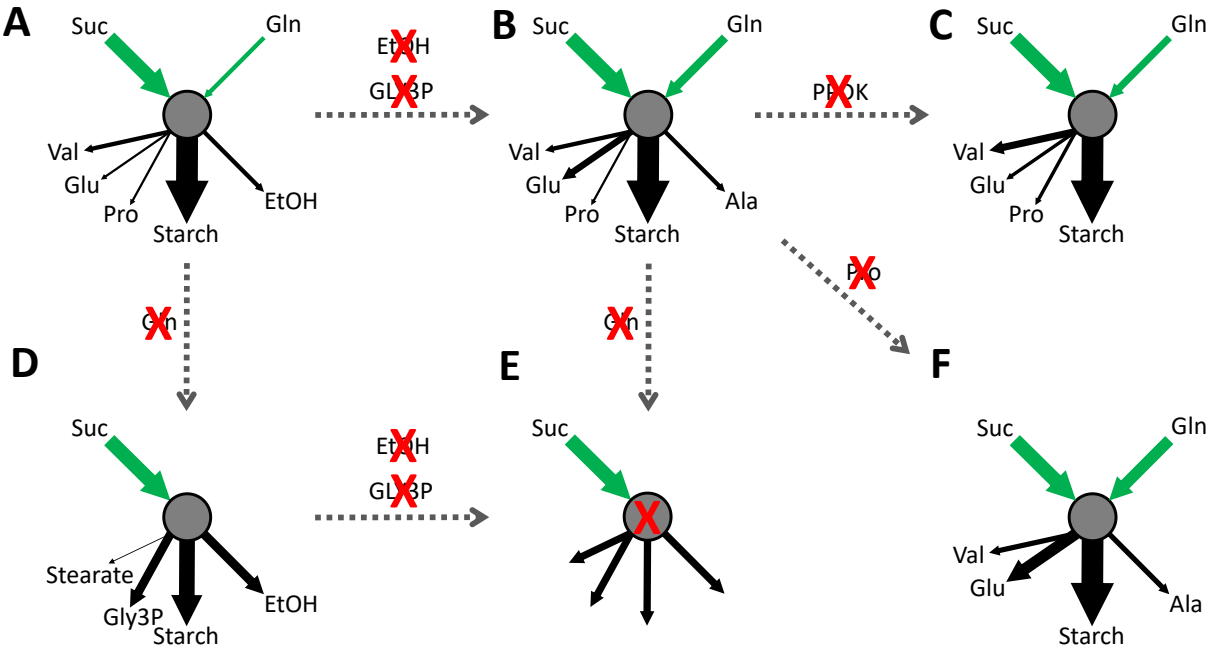


Figure 4.7. Iterative methodology employed in restricting specific metabolite excretion in FBA simulations. The figure depicts the constraints placed on the FBA and FVA simulations. We adopted an iterative method to constrain biologically meaningful FBA solutions. The circles represent the system boundaries of endosperms. Green arrows going to the circle represent nutrients while black arrows going out of the circle represent excretion fluxes. The arrow sizes are based off of the size of the fluxes predicted by FBA. Arrows between different panels, with either metabolite or enzyme name with an "X" represents the additional constraint placed on the model. All abbreviations of metabolites are defined in the List of Abbreviations. A, nutrients and excreted metabolites predicted by FBA of WT endosperm. B, nutrients and excreted metabolites predicted by FBA of WT endosperm after blocking excretion of glycerol-3-phosphate and ethanol. C, nutrients and excreted metabolites predicted by FBA of PPDK deficient Tx endosperm after blocking excretion of glycerol-3-phosphate and ethanol. D, nutrients and excreted metabolites predicted by FBA of WT endosperm, without Gln uptake. E, no solution in FBA of Tx endosperm, without Gln uptake. F, nutrients and excreted metabolites predicted by FBA of WT endosperm after blocking excretion of glycerol-3-phosphate, ethanol and Pro.

4.4 Discussion

We computationally investigated grain filling in maize endosperms by placing physiologically relevant constraints on the reconstructed GSM. Each optimal solution from a simulation is an extreme pathway (Klamt and Stelling, 2003), and representation of optimal metabolism under the physiological constraints simulated by the constraints. Plant metabolic networks rarely operate optimally. Hence the results must be analyzed carefully in the context of known biochemical evidence and constraints posed by the nutrient microenvironment.

4.4.1 An iterative approach to genome-scale FBA enables faster hypothesis development

We adopted an iterative method involving the following steps, not necessarily in the order mentioned, in our investigations: (i) FBA with a minimal set of constraints and analysis of extreme pathways; (ii) removal of biologically unrealistic predictions from the solution space and repetition of FBA, and extreme pathway analysis, with the updated constraints; (iii) FVA to assess the variability of the initial predictions and to identify factors essential to optimality; (iv) identification of the metabolic reasons behind the essential factors identified in the previous step. We employed (i) to (ii) to remove the biologically unrealistic predictions of ethanol and GLY3P excretion and investigate WT endosperm metabolism (**Figure 4.4A**, **Figure 4.4B**). The same procedure (steps (i) to (ii)), with additional constraint of PPDK blockage, were employed for predicting Tx endosperm metabolism during starch filling (**Figure 4.4A, Figure 4.4B, Figure 4.4C**). The third step was essential to identification of metabolic rigidity of Gln uptake and amino acid fermentation. The last step identified the importance of Pro excretion on minimizing Gln uptake. Additionally, employing (i) to (iii) without Gln uptake the simulations did not converge, unless the biologically unrealistic excretions of either ethanol or GLY3P were restored (**Figure 4.4D**). We stress that these observations can be identified by repeatedly performing FBA. However, performing FBA *ad infinitum* is time-consuming and an iterative methodology speeds up the workflow significantly. The altered amino acid profile of Tx endosperm, predicted by FBA, agrees with our experimental data (**Figure 4.3**).

4.4.2 Essentiality of Gln uptake in the simulations is necessitated by reductant cofactor utilization

Glycolysis in endosperm is a high flux pathway (Rolletschek et al., 2011). However, glycolysis is inefficient in generating ATP, in comparison to TCA cycle. Glycolysis, in addition to ATP, generates reductant cofactors, which must be consumed for sustained hexose metabolism. Regeneration of these reductant cofactors requires a terminal electron acceptor. Biosynthesis of Pro, Val and Glu from Gln expends the reductant cofactors, allowing sustained glycolytic flux. The enzymes glutamate synthase and acetolactate synthase, which are essential to the biosynthesis pathways of Glu and Val respectively, are highly expressed in the endosperm during starch filling (Sodek and da Silva, 1977; Muhitch, 1988), which supports our FBA predictions. The synthesis of Glu and Pro does not need any C derived from Suc, but, consumes reductant cofactors. Val biosynthesis utilizes Pyr, which the endosperm needs to expend, and consumes reductant cofactors. Pro biosynthesis efficiently consumes more reductant than any other amino acid without requiring extra carbon. Additionally, blocking these amino acids resulted in the excretion of other amino acids and a small decrease in starch synthesis. Most notably, blocking Val resulted in the excretion of Leu, which is seen in high quantities within α-zeins. Until now, concurrence of starch and amino acid biosynthesis in endosperm (Sabelli and Larkins, 2009) has largely been attributed to storage protein synthesis. To the best of our knowledge, amino acid metabolism and regeneration of cofactors as a metabolic bottleneck during starch filling has not been suggested earlier. These predictions are supported by our experimental observations (**Figure 4.3**).

4.4.3 Gln-supported metabolism explains spatiotemporal variations in endosperm metabolite profile

The amount of dissolved O_2 in the endosperm drops dramatically upon diffusion through the pericarp (Rolletschek et al., 2005; Borisjuk and Rolletschek, 2009). This diffused O_2 is available to a small outermost section of the endosperm (Rolletschek et al., 2005). Thus, some activity of TCA cycle in this outermost section of endosperm cannot be ruled out. A recent experimental study has reported significantly higher amounts of free amino acids, including Glu, Gln, Pro and Val in the more hypoxic endosperm center, compared to endosperm periphery (Gayral et al., 2017). This spatial variation can be explained by a greater reliance of the significantly more hypoxic endosperm center on Gln uptake and subsequent biosynthesis of other amino acids, to expend the reductant cofactors. Similar response to hypoxia has been observed in maize roots with the accumulation of ethanol, lactate, Ala and GABA during water-logging (Antonio et al., 2015). Additionally, another study employing integrated transcriptomic, metabolomics and proteomics study had reported an increase in amino acid metabolism from between 10 DAP and 14, and 21 DAP (Prioul Jean Louis et al., 2008). This is likely because of the progressive increases in endosperm size and density with an increase in storage product accumulation, which will limit O_2 diffusion further. Thus, the reliance on Gln uptake and amino acid synthesis increases as grain filling progresses. Ala is synthesized from glycolytic Pyr, with nitrogen derived from Gln. Elevated Ala accumulation in the endosperm center has been reported earlier (Rolletschek et al., 2011). These experimental observations fully support our hypothesis of Gln metabolism being a major factor in accumulation of storage products.

4.4.4 PPDK permits redundancy of ADP-Glc synthesis and subsequent PPi recycling in cytosol and plastid

In maize endosperm, 85% to 95% of total AGPase activity was cytosolic (Denyer et al., 1996). During ADP-Glc synthesis, PPi is generated at the cost of ATP. The simulations predicted that recycling of PPi by cytosolic *PFP* and UDP–glucose pyrophosphorylase (UDPgp) is essential to optimal starch synthesis (**Figure 4.5, Figure 4.6**). Recycling of PPi by UDPgp involves a phosphorylation cycle that includes *sus1.* Earlier, *sus1* has been identified to conserve ATP (Rolletschek et al., 2005). Maize genome does not contain a cytosolic pyrophosphatase to breakdown PPi, which could potentially inhibit AGPase (Boehlein et al., 2010). Recycling of PPi decreases the ATP demand for phosphorylation. The essentiality of *PFP*, as predicted by our FVA simulations, has not been suggested before. In the plastid, PPDK and inorganic pyrophosphatase (PPase) are the only PPi-consuming reactions available. However, PPDK produces an ATP molecule through PPi consumption, making it a desirable reaction for decreasing the ATP demand during plastidic ADP-Glc production.

In Tx endosperm, plastidic ADP-Glc synthesis is predicted to be suboptimal during maximal starch synthesis. When simulating the absence of plastidic ADP-Glc synthesis in WT endosperms, PPDK became inactive and starch synthesis was equal to the amount predicted in Tx endosperms. The decrease in starch in our simulations due to the loss of PPDK and plastidic AGPase mirrored experimental results that compared starch dry weights of WT wheat endosperms and wheat endosperms deficient in plastidic AGPase (Huang et al., 2014). When compared to previous experimental observations, these simulations suggest that PPDK is responsible for ADP-Glc synthesis in the plastid during

grain filling periods in maize endosperm. Our FVA results also revealed that enzymatic activity in the Tx endosperm is less flexible than WT endosperms (**Figure 4.5, Figure 4.6**), further suggesting that PPDK provides maize endosperm with metabolic versatility during starch synthesis.

4.4.5 Predictive modeling with limited constraints: a best-case scenario in a worst-case scenario

Plants, being sessile organisms, adjust their metabolism in response to environmental challenges. Thus, the steady-state optimal flux predictions capture only a snapshot of the metabolic extremities. FBA predicted optimal solution can be very sensitive to constraints imposed. Ideally, all constraints should be based on experimental evidence. Without this information, questions posed to FBA must be framed carefully to get meaningful qualitative answers.

The constraints Suc and Gln being the sole nutrients source of endosperms, in our simulations, is very restrictive. Multiple amino acids are transported to developing maize kernels in smaller quantities as compared to Gln (Sodek and da Silva, 1977; Arruda and Da Silva, 1979). Ala, Glu, Pro and Val excretions is a result of imposed nutrient constraints, structure of the metabolic network and the applied objective function. A change in imposed constraints or objective function would most likely lead to other amino acids being excreted. Without constraining the network further, either by selecting a subset of reactions of the GSM or by incorporating more experimentally measure constraints, extraction of meaningful qualitative information from FBA performed with all amino acids as endosperm nutrients (Arruda and Da Silva, 1979) becomes nearly impossible. Thus, an iterative and selective workflow employing a minimalistic set of constraints aids efficient hypothesis development for experiment design.

4.5 Conclusion

We developed a genome-scale model of maize and used flux balance modeling to seek answers to energetics grain filling. The predictions have led to the hypothesis that amino acid fermentation is a way of dealing with hypoxia in the endosperm. Additionally, we have hypothesized that redundancy in ATP generating pathways alleviated the dependency on glycolytic ATP generation by enabling recycling of PPi. Further investigations with targeted gene knockouts will be necessary to validate these predictions.

Chapter 5 : Integration of transcriptome data with a genome-scale stoichiometric model of maize for investigation of storage protein accumulation in endosperm

This chapter is part of a manuscript to under preparation. Copyright maybe transferred to the publisher in the future.

Navadeep Boruah, Brendon K. Cox and Ganesh Sriram*

NB and GS conceived this project. NB developed the mathematical formulation. NB and BKC performed the analyses. NB wrote the draft.

Department of Chemical and Biomolecular Engineering (N. B., B. K. C., G. S.), University of Maryland, College Park, MD 20742, United States of America

5.1 Background

The nutrient content of grains is an important aspect of its quality. The endosperm in cereals is the primary nutrient storage site of the kernels. The nutrient reserve in endosperm is necessary for germination of the embryo and early growth. While starch is the major biomass component in maize endosperms, it also contains about 10% protein (Gibbon and Larkins, 2005; Zhang et al., 2015). Prolamins are the principal storage proteins in cereals (Gibbon and Larkins, 2005; Zhang et al., 2015). Zeins are the prolamins in maize endosperms, contributing to over 70% of total endosperm proteins (Gibbon and Larkins, 2005; Guo et al., 2013). Zeins are poor quality proteins due to lack of essential amino acids Lys and Trp (Gibbon and Larkins, 2005; Zhang et al., 2015). Alterations of zein profile to yield a higher quality kernel is a nontrivial task. In addition to N storage, zeins in endosperm contribute to physical properties of the endosperm. Naturally occurring *opaque2* (*o2*) mutant kernels with lower zein but higher Lys content have poor vitreousness (Guo et al., 2013). Kernel vitreousness is necessary for resistance to insect and fungal damage, and stronger mechanical properties. The lack of hard endosperm in the mutant makes it agronomically unsuitable. Conventional plant breeding over decades have led to the development of quality protein maize (QPM), a hard kernel mutant containing higher levels of Lys and Trp (Guo et al., 2013). However, the reasons behind the lack of the two essential amino acids, and the predominance of certain amino acids in zeins (**Figure 5.1**) remain unclear.

Zein synthesis in endosperm long been considered a phenomenon associated with starch accumulation during grain filling. Zeins are classified into α (19 and 22 kDa), β (15 kDa), γ (50, 27 and 16 kDa) and δ (10 and 18 kDa) subfamilies depending upon solubility (Zhang et al., 2015). The α-zeins, containing high relative amount of Glu (25%), Leu (20%) and Ala (11%), contribute to 60-70% of total zein content (Gianazza et al., 1977; Thompson and Larkins, 1994; Zhang et al., 2015). The γ-zeins with unusually high Pro content (25%) account for 20-25% of total zeins (Thompson and Larkins, 1994; Zhang et al., 2015). Additionally, the spatiotemporal profiles of expression of zein genes and, thus the proteins themselves, are not identical (Thompson and Larkins, 1994). Accumulation of γ-zeins starts earlier than the other zein classes. The vitreous periphery of the kernel

contains more zeins than the opaque kernel center, α-zeins being predominant in endosperm center (Thompson and Larkins, 1994).

Oxygen availability and associated metabolism has been known to play a key role in this dynamics of kernel metabolism (Prioul Jean Louis et al., 2008). Changing transcriptome and proteome expression profiles (Prioul Jean Louis et al., 2008). and kernel nutrient profile (Arruda and Da Silva, 1979) throughout the duration of grain filling have been observed in experimental studies. It is unclear if dynamic metabolic landscape of kernel metabolism is a factor behind the deposition of specific zeins. The analyses of the previous chapter had suggested elevated synthesis of multiple amino acids (Ala, Glu and Pro) that are predominant in different zeins as a way of regenerating the glycolytic reductant cofactors. However, the factors determining deposition of zeins remained unclear.

We extended that work here by if any correlation exists between zein accumulation and metabolic network structure under a given nutrient constraints. Here, we integrated publicly available time-resolved transcriptome datasets with our maize genome-scale model by using a bilevel optimization based computational. Specifically, we sought explanations of the following metabolic phenomena observed during grain filling: (i) why do different zein proteins get deposited at different time stages of grain filling; (ii) what necessitates shifts in gene expression profiles with a progress in grain-filling; and (iii) how does flux variability changes over time during grain filling. We discuss the computational framework and the results from the analyses in the following sections.

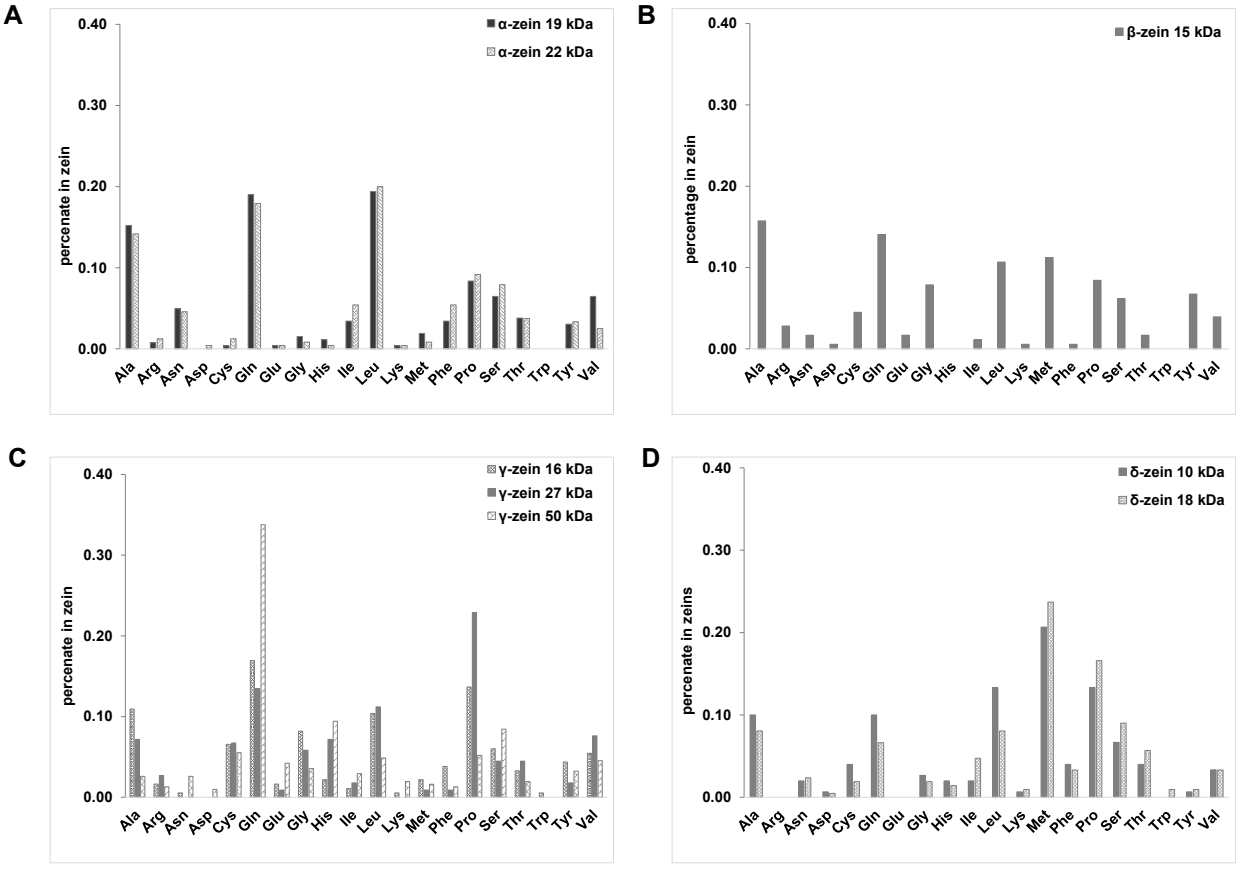


Figure 5.1. Distribution of all proteinogenic amino acids in zein proteins. We extracted the amino acid sequences of all known maize prolamins from the UniProt database and determined their amino acid distributions. A, distribution of proteinogenic amino acids in α-zeins. B, distribution of proteinogenic amino acids in β-zeins. C, distribution of proteinogenic amino acids in γ-zeins. D, distribution of proteinogenic amino acids in δ-zeins.

5.2 Integration of Transcriptome Datasets with GSMs: An Overview of Past Research

Genome-scale models of metabolism with integration of omics data is commonly used method of constraining the large solution space typically associated with FBA (Reed, 2012). Of all omics technologies, gene expression is the most widely used as it is amenable to amplification. With decreasing costs of NGS sequencing, it is now possible to generate large datasets much more readily. In theory, metabolic flux depends upon enzyme levels, and hence on transcript abundance (Shi and Schwender, 2016a). However, constraining with transcriptome data does not account for post-transcriptional and post-translational modifications (Schwender et al., 2014). Generally, transcript abundances have low correlations with metabolic fluxes (Kim and Lun, 2014; Shi and Schwender, 2016a) And hence they alone cannot be used to estimate metabolic fluxes in central carbon metabolism (Schwender et al., 2014).The fidelity of predictions improves with the increase in the number of experimental constraints, such as proteomics data, as additional constraints (Shi and Schwender, 2016a) or regulatory information at transcriptional, translational and/or post-translational stages (Machado and Herrgård, 2014; Saha et al., 2014). Scores of computational frameworks for integrating genome-wide transcriptomics data with GSMs have been developed in the last 15 years. However, the experimental information necessary as input to many of these methods is either not available or expensive to obtain, thereby limiting application in plant systems. Only a handful of articles (Töpfer et al., 2012; Töpfer et al., 2013; Simons et al., 2014bSimons et al., 2014a; Dal'Molin et al., 2016; Lakshmanan et al., 2016; Scheunemann et al., 2018) integrating transcriptomics data with GSMs have been published so far. For example, constraints like biomass equation or nutrient uptake for specialized plant cells in developmental stages can be nearly impossible to obtain. Addressing these challenges is a necessity while integrating transcriptomics data with plant GSMs.

The computational frameworks developed for transcriptome integration with GSMs vary widely in terms of dataset requirement, biological assumptions and objective behind the formulations. The methods have broadly been grouped in the following categories (Kim and Lun, 2014; Machado and Herrgård, 2014): (i) ones that require multiple; (ii) one that require arbitrary thresholds; (iii) validation of fluxes against experimentally measured fluxes; (iv) discrete vs continuous levels; (v) absolute vs relative expression. The selection of a subset of reactions with transcriptome data have been involved using Boolean logic (Åkesson et al., 2004), baseline transcript levels (Becker and Palsson, 2008), maximizing consistency with highly expressed and genes (Shlomi et al., 2008), minimizing flux-transcripts violations (van Berlo et al., 2011), minimizing predicted to flux –gene expression differences (Lee et al., 2012), mapping expression levels to flux bounds (Colijn et al., 2009), etc. Additional layers of information like regulatory information of gene regulation (Chandrasekaran and Price, 2010), metabolite turnover

(Schmidt et al., 2013), p-values of gene expression (Jensen and Papin, 2011), enzyme-metabolite interaction (Moxley et al., 2009). These methods have seen applications in mammalian/cancer cells (Shlomi et al., 2008; Agren et al., 2012; Opdam et al., 2017) and microbes (Maia et al., 2016; Kim et(Song et al., 2014) al., 2018) alike. Recently tissue specific models for Arabidopsis have been generated as well, which has been validated with proteomics data (Scheunemann et al., 2018). Comparative analyses of some of these transcriptome integration frameworks have been reported earlier (Kim and Lun, 2014; Machado and Herrgård, 2014).

Metabolic fluxes estimated with GSM integrated with transcriptome data generally has low fidelity. No method that predicts flux accurately under all conditions, even in bacteria (Machado and Herrgård, 2014). In fact, growth yield and parsimony alone provided better estimates of metabolic fluxes than transcriptomes of most experimental sets (Machado and Herrgård, 2014). Often, equally optimal alternate solutions also coexist (Shi and Schwender, 2016a). Recent studies have shown that model extraction method (Opdam et al., 2017), and thus model structure (Yuan et al., 2016a), has the largest impact on model accuracy. Additionally, cellular energy maintenance costs (Cheung et al., 2013), addition of ^{13}C-MFA data (Shi and Schwender, 2016a) and addition of different norm minimization (Shi and Schwender, 2016a) also impact the quality of solutions obtained. Central metabolic fluxes were largely insensitive to perturbations in biomass compositional variations, but were affected by model structure (Yuan et al., 2016a). The proteome profile is more stable than mRNA profile (Walley et al., 2013). Hence, proteome might be a better indicator of metabolic activity than transcriptome. Obtaining proteome dataset is significantly more challenging. Hence, we focus on gaining meaningful qualitative information about endosperm metabolism over actual flux values.

5.3 Materials and Methods

5.3.1 Model Curation

We employed the maize GSM developed for our analyses in Chapter 4, with the addition of reactions for zein synthesis and excretion for our analyses of this chapter. We obtained the zein sequences from the UniProt database and created zein synthesis reactions with the molar distributions of amino acids in the zein sequences. These zein synthesis reactions, along with zein secretion reactions, were added to the model. To reduce computation time with our bilevel optimization framework, we created a condensed version of our GSM by picking only those reactions that could carry nonzero fluxes with free flow of a wide range of biomass constituents, inorganic nutrients and common carbohydrate sources across the system boundary. The condensed model contained 1213 reactions, 841 of which were metabolic reactions.

5.3.2 Transcriptome Datasets

We employed the datasets generated by Chen et al. with RNA-seq (Illumina) to analyze spatiotemporal variations in B73 maize transcriptome, to investigate zein accumulation during grain-filling (Chen et al., 2014). This dataset contains 17 endosperm samples collected at different time points of kernel development, out of a total of 78 samples obtained from maize embryo, endosperm, whole seed, shoots, root, leaf, shoot epical

meristem, ear, tassel, cob, silk, anthers, ovule and pollen. The inclusion of datasets from such a wide variety of tissues at different time points provided us with sufficient number of discrete expression levels for estimation of the three parameters of the logistics equation.

5.3.3 Curation of gene-to-reaction association

The data files obtained from the Maizecyc and Corncyc databases at times contained conflicting information on gene-protein-reaction associations. The differences between these two databases appear due of differences in annotation pipelines (Walsh et al., 2016). If a gene appeared in both databases, but was matched to different reactions, the annotation of CornCyc was chosen over that of MaizeCyc due to generally higher accuracy of CornCyc (Walsh et al., 2016). Often genes were found to be associated with multiple reactions. Promiscuous enzymes are a common trope in plant secondary metabolism, but enzymes in primary metabolism tend to be conservative (Shi and Schwender, 2016b). Genes associated with more than one reactions were thoroughly inspected, which often turned out to be specific reactions with a common functional group. Else, we picked the annotation of CornCyc database. This procedure was performed for all reactions in MaizeModel.

5.3.4 Challenges, Limitations of Constraints/Datasets and Assumptions

A major objective behind the development of our computational framework was to address the challenges encountered in predictive modeling of plant tissues. The quality of predictions from optimization based predictive modeling frameworks is reliant the constraints and objective function employed in the analyses. These constraints are often obtained from experimental measurements. However, time resolved experimental data on complete nutrient profile and biomass composition in developing plant tissues is not widely reported in the literature. Hence, a major goal behind the development of our computational framework was addressing these challenges, posed by a lack of data on nutrient and biomass profiles, by harnessing the relative ease of large-scale transcriptome dataset generation. A major factor driving the development of our computational framework was the inability of the previously reported frameworks to incorporate a biological objective at a developmental stage along with selection of a set of reactions without any *a priori* assumption on reaction flux bound or threshold expression level for categorization of a reaction or a reference state.

We addressed the lack of nutrient measurements B73 maize endosperm by using amino acid composition of a high lysine maize ear peduncle. This data, while not obtained from the same maize line, contain the information on the changing nutrient profile during grain filling. This is sufficient for our qualitative analysis of the impact of nutrient profile on zein synthesis. If accurate estimation of metabolic fluxes is an objective then incorporation of nutrient measurements B73 maize alone will most likely not be sufficient as transcript abundances alone cannot be employed to predict metabolic fluxes in central carbon metabolism of plant tissues (Schwender et al., 2014).

As the temporal profile of storage compounds in endosperm changes drastically over time rendering the cellular objective unclear, we employ an iterative approach in which a biological objective is selected *a priori* and then the results are analyzed against reported

data. This selection of objective is based on known biochemical knowledge of storage product synthesis. For example, starch and zeins account for over 90% of endosperm biomass, which much of this synthesis taking place between 12 DAP-28 DAP. Hence, synthesis of this storage products can be a suitable cellular objective once grain filling progresses. We stress that depending upon the biological information available, and the tissue of interest, our computational framework might require incorporation of different constraints to reflect the qualitative cellular objective.

5.3.5 Assumptions in Our Framework

The major assumption in our framework is that for each enzyme coding genes there is data encompassing on all possible levels (low, high, medium etc.) of the gene's expression in the form of transcript abundances. We fit the data for each gene with a logistics equation with three parameters. The data was used to estimate a parameter that indicate the probability of the gene coding for an active enzyme. To fit the data, we sort the dataset from the lowest to highest expression level on equidistant points and associate the lowest measured value of a gene's expression with a probability of zero and the highest level with a probability of one (**Figure 5.2**). We note that this nonlinear mapping of gene expression of probability on equidistance points is arbitrary, as no such experimentally determined probability exists. Once the parameters are estimated, they can be employed, with a specific transcript levels of a gene, to calculate the probability parameter for that gene. The estimated parameters of logistics equation, and the probability mapping is robust if the data fitting is done with sufficient data points covering all possible ranges of a gene's expression. We group the reactions in the network as high, moderate or low probability reactions based on the probability, and α which is a user chosen parameter. We note that gene expression does not necessarily relate to fluxes, and hence none of the low probability reactions are restricted from carrying fluxes in our framework. We discuss the logistics equation and the computational framework in the following sections.

5.3.6 Logistics Equation

We employed the following three parameter logistics equation for estimation of the probabilities. The three parameters $K, r \text{ and } X_0$ specify the maximum asymptotic value, the slope at inflexion point and the position of inflexion point respectively. The probability p_{jk} for gene k of reaction j can be defined as,

$$p_{jk} = \frac{K}{1 + \quad}$$

We transform the discrete expression levels, after sorting in the increasing order, to confine to the range $[0,1]$ with the following equation.

$$X = \frac{(x)-\min\{\log(x)\}}{\max\{\log(x)-\min\{\log(x)\}\}}$$

We assume that the transformed X variables correspond to incrementally increasing probabilities (let's say p_y) of the reaction carrying nonzero metabolic flux. Once the variables are transformed, the parameters $K, r \text{ and } X_0$ for each gene are estimated by minimizing the sum of squares of the differences between the probabilities calculated with the estimated parameters (p_{jk}) and the assumed probabilities (p_y). With sufficiently large number of points, the estimated probabilities are quite robust.

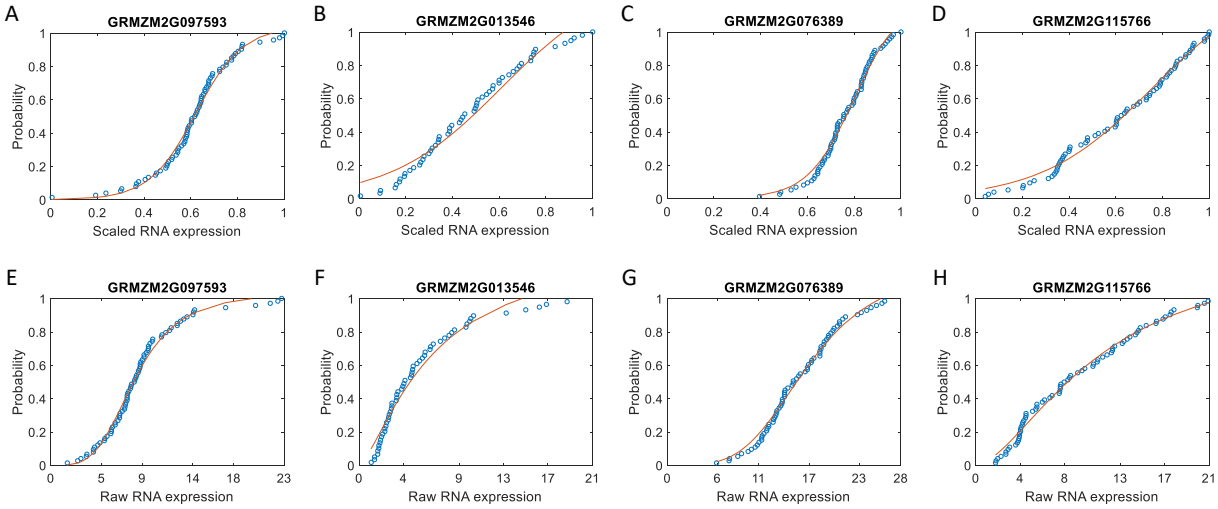


Figure 5.2. Plot of estimated probabilities p_{jk} against the assumed probabilities p_y.
The blue circles in each plot indicate the actual discrete data of expression levels, while the red lines show the mapping of the probabilities with the estimated parameters. Each gene has its unique shape, and hence logistics equation parameters. Panels A-D show the probability mapping with scaled expression levels, with scaled expression levels on x-axis and the probabilities on y-axis, for four randomly picked genes. Panels E-H show the expression levels without scaling for the genes of A-D, with actual expression levels on x-axis and the probabilities on y-axis.

Based on a user defined parameter α we classify a reaction as high or low probability reaction. In the outer objective function, we maximize the total number of high probability reactions carrying fluxes. The inner objective function maximizes a hypothesized objective function, which is the basis of the investigation.

5.3.7 Computational Framework

5.3.7.1 Formulation of bilevel optimization and reaction classification

We employ a simplified bilevel formulation with the transcriptome matching function as the outer objective function, and starch maximization as the inner objective function.

$$\max \; z_{match} = \sum_{\forall j \in H} (d_j^{rev} + d_j^{for}) p_j - \sum_{\forall j \in L} (d_j^{rev} + d_j^{for})(1 - p_j)$$

s.t.

$\max z_{inner}$

s.t.

$S_{ij}v_j = b_i \quad \forall j \quad \rightarrow \quad m_i \in R$

$v_j \Delta g_j \le 0 \quad \forall j \notin \{\text{uptake, secrete}\} \quad \rightarrow \quad e \ge 0$

$-v_j \le -lb_j d_j^{rev} \quad \forall j \notin \{\text{uptake, secrete}\} \quad \rightarrow \quad \mathrm{u}_j^{LB} \ge 0$

$v_j \le ub_j d_j^{for} \quad \forall j \notin \{\text{uptake, secrete}\} \quad \rightarrow \quad \mathrm{u}_j^{UB} \ge 0$

$-v_j \le -v_j^{lb} \quad \forall j \in \{\text{uptake}\} \quad \rightarrow \quad \mathrm{u}_j^{LB} \ge 0$

$v_j \le v_j^{ub} \quad \forall j \in \{\text{uptake}\} \quad \rightarrow \quad \mathrm{u}_j^{UB} \ge 0$

$-v_j \le 0 \quad \forall j \in \{\text{secrete}\} \quad \rightarrow \quad \mathrm{u}_{j \in \text{secrete}}^{LB}$

$v_j \le v_j^{ub} \quad \forall j \in \{\text{secrete}\} \quad \rightarrow \quad \mathrm{u}_{j \in \text{secrete}}^{UB}$

$d_j^{rev} + d_j^{for} \le 1,\ d_j^{rev}, d_j^{for} = \{0,1\}$

$0 \le \alpha \le 0.5$ **(user specified)**

$p_j = f(p_{jk}),\ p_{jk} = f_{jk}(x_{jk}),\ x_{jk,\min} \le x_{jk} \le x_{jk,\max}$

We classified each reaction as $\mathrm{j} \in$, $\mathrm{j} \in$ or $\mathrm{j} \notin$ depending upon the user chosen parameter α . This information is stored in a vector $\mathrm{ii}_{\text{class}} \in$.

$\mathrm{ii}_{\mathrm{j}}^{\text{class}} = \rightarrow \in$

$\mathrm{ii}_{\mathrm{j}}^{\text{class}} = \rightarrow \notin$

$\mathrm{ii}_{\mathrm{j}}^{\text{class}} = \rightarrow \in$

We used the following classification algorithm to classify reaction j based on all genes jk of that reaction.

Step 1) Pick a value of say $\alpha = 0.25$. Ideally you want it to be as small as possible.
Step 2) Identify all the genes of reaction j such that $p_{jk} \in]\alpha, 1]$ and the mean $\overline{p_{jk}} \forall p_{jk} \in]\alpha, 1]$.
Step 3) If $\overline{p_{jk}} > 0.5 \rightarrow j \in H,\ p_j = \max\{p_{jk}\}$ and $\mathrm{ii}_{\mathrm{j}}^{\text{class}} =$.
Step 4) If $\forall k : \overline{p_{jk}} \in [0, \alpha] \rightarrow j \in L,\ p_j = \min\{p_{jk}\}$ and $\mathrm{ii}_{\mathrm{j}}^{\text{class}} =$.
Step 5) If steps (3)-(4) contradicts classification of reaction $\mathrm{j} \rightarrow \mathrm{j} \notin$ and $\mathrm{ii}_{\mathrm{j}}^{\text{class}} =$.
Step 6) Identify all reactions $\mathrm{j} \notin$.

We stress that the prior to using the bilevel optimization framework, the inner problem should be solved to ascertain feasibility. If the inner problem is infeasible, then the bilevel problem is infeasible as well. The parameter α can be used to "tune" the results by adopting an iterative approach in which the bilevel problem is updated by changing α , if the predictions for the previous value of α did not match experimental data. This might also be useful if the predictions are compared against ^{13}C-MFA derived flux data as predictions from the bilevel frameworks will be different for different values of α . Additionally, care must be taken to ensure that the inner objective value z_{inner} is nonzero. If the solution of the bilevel problem returns a value of zero for z_{inner} then the problem can be strengthened by adding a constraint of nonzero value of z_{inner} .

5.3.7.2 Conversion to a single level problem

We employed the duality theory of linear programing to convert the bilevel optimization problem to a single level problem. For the inner problem of starch maximization, the bilevel problem can be written as following:

$$\max\ z_{match} = \sum_{\forall j \in H} (d_j^{rev} + d_j^{for}) p_j - \sum_{\forall j \in L} (d_j^{rev} + d_j^{for})(1 - p_j)$$

s.t.

$$\max z_{inner} = v_{\forall j \in \{starch\}}$$

s.t.

$$S_{ij} v_j = 0 \qquad \forall i \qquad \rightarrow \qquad m_i \in R$$

$$-v_j \le -LB_j d_j^{rev} \qquad \forall j \notin \{\text{uptake, secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j \notin \{\text{uptake, secrete}\}}^{LB} \ge 0$$

$$v_j \le UB_j d_j^{for} \qquad \forall j \notin \{\text{uptake, secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j \notin \{\text{uptake, secrete}\}}^{UB} \ge 0$$

$$-v_j \le -v_j^{lb} \qquad \forall j \in \{\text{uptake}\} \qquad \rightarrow \qquad \mathrm{u}_{j \in \{\text{uptake}\}}^{LB} \ge 0$$

$$v_j \le v_j^{ub} \qquad \forall j \in \{\text{uptake}\} \qquad \rightarrow \qquad \mathrm{u}_{j \in \{\text{uptake}\}}^{UB} \ge 0$$

$$-v_j \le 0 \qquad \forall j \in \{\text{secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j \in \{\text{secrete}\}}^{LB} \ge 0$$

$$v_j \le v_j^{ub} \qquad \forall j \in \{\text{secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j \in \{\text{secrete}\}}^{UB} \ge 0$$

$$d_j^{rev} + d_j^{for} \le 1,\ d_j^{rev}, d_j^{for} = \{0,1\}$$

The dual of the inner optimization problem can be written as following:

$$\min z_{inner,dual} = -(LB_j d_j^{rev} \mathrm{u}_j^{LB})_{j \notin \{\text{uptake, secrete}\}} + (UB_j d_j^{for} \mathrm{u}_j^{UB})_{j \notin \{\text{uptake, secrete}\}}$$

$$-v_{j \in \{uptake\}}^{LB} \mathrm{u}_{j \in \{uptake\}}^{LB} + v_{j \in \{uptake\}}^{UB} \mathrm{u}_{j \in \{uptake\}}^{UB} + v_{j \in \{\mathrm{sec}\,rete\}}^{UB} \mathrm{u}_{j \in \{\mathrm{sec}\,rete\}}^{UB} \qquad \forall j$$

$$S_{ji} m_i + \mathrm{u}_j^{ub} - \mathrm{u}_j^{lb} = 0 \qquad \forall j - \{starch\}$$

$$S_{ji} m_i + \mathrm{u}_j^{ub} - \mathrm{u}_j^{lb} = 1 \qquad \forall j \in \{\text{starch}\}$$

$$\mathrm{u}_j^{LB},\ \mathrm{u}_j^{uB} \ge 0 \qquad \forall j$$

$$\mathrm{u}_j^{LB} \le \mathrm{u}_j^{LB,\max} \qquad \forall j$$

$$\mathrm{u}_j^{UB} \le \mathrm{u}_j^{UB,\max} \qquad \forall j$$

We invoke the strong duality theory of linear programing at optimality which demands that at optimality the values of the primal and dual objectives are equal. Thus, we can rewrite the bilevel optimization problem as a single level problem.

$$\max\ z_{match} = \sum_{\forall j \in H} (d_j^{rev} + d_j^{for}) p_j - \sum_{\forall j \in L} (d_j^{rev} + d_j^{for})(1 - p_j)$$

s.t.

$$S_{ij}v_j = 0 \qquad \forall i \qquad \rightarrow \qquad m_i \in R$$

$$-v_j \le -LB_j d_j^{rev} \qquad \forall j \notin \{\text{uptake, secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j\notin\{\text{uptake, secrete}\}}^{LB} \ge 0$$

$$v_j \le UB_j d_j^{for} \qquad \forall j \notin \{\text{uptake, secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j\notin\{\text{uptake, secrete}\}}^{UB} \ge 0$$

$$-v_j \le -v_j^{lb} \qquad \forall j \in \{\text{uptake}\} \qquad \rightarrow \qquad \mathrm{u}_{j\in\{\text{uptake}\}}^{LB} \ge 0$$

$$v_j \le v_j^{ub} \qquad \forall j \in \{\text{uptake}\} \qquad \rightarrow \qquad \mathrm{u}_{j\in\{\text{uptake}\}}^{UB} \ge 0$$

$$-v_j \le 0 \qquad \forall j \in \{\text{secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j\in\{\text{secrete}\}}^{LB} \ge 0$$

$$v_j \le v_j^{ub} \qquad \forall j \in \{\text{secrete}\} \qquad \rightarrow \qquad \mathrm{u}_{j\in\{\text{secrete}\}}^{UB} \ge 0$$

$$d_j^{rev} + d_j^{for} \le 1,\ d_j^{rev}, d_j^{for} = \{0,1\}$$

$$v_{\forall j\in\{starch\}} = -(LB_j d_j^{rev} \mathrm{u}_j^{LB})_{j\notin\{\text{uptake, secrete}\}} + (UB_j d_j^{for} \mathrm{u}_j^{UB})_{j\notin\{\text{uptake, secrete}\}}$$

$$-v_{j\in\{uptake\}}^{LB}\mathrm{u}_{j\in\{uptake\}}^{LB} + v_{j\in\{uptake\}}^{UB}\mathrm{u}_{j\in\{uptake\}}^{UB} + v_{j\in\{\sec rete\}}^{UB}\mathrm{u}_{j\in\{\sec rete\}}^{UB} \qquad \forall j$$

$$S_{ji}m_i + \mathrm{u}_j^{ub} - \mathrm{u}_j^{lb} = 0 \qquad \forall j - \{starch\}$$

$$S_{ji}m_i + \mathrm{u}_j^{ub} - \mathrm{u}_j^{lb} = 1 \qquad \forall j \in \{\text{starch}\}$$

$$\mathrm{u}_j^{LB},\ \mathrm{u}_j^{uB} \ge 0 \qquad \forall j$$

$$\mathrm{u}_j^{LB} \le \mathrm{u}_j^{LB,\max} \qquad \forall j$$

$$\mathrm{u}_j^{UB} \le \mathrm{u}_j^{UB,\max} \qquad \forall j$$

All the terms in the dual objective function are bilinear. Thus, the bilevel programing problem is rendered to a single level mixed integer nonlinear programing problem.

5.3.8 Software

All the simulations in this chapter were carried out with GAMS 24.7.1 with the solver CPLEX 12.

5.4 Results

5.4.1 Single level optimization framework cannot predict zein synthesis

We investigated if a single level optimization framework employing gene expression, without any *a priori* assumption of cellular objective, is sufficient to accurately predict onset of specific zeins during grain filling. We employed maximization of the transcriptome matching function as the objective and constrained the model with experimentally measured amino acids distributions in vascular sap of maize peduncle (Arruda and Da Silva, 1979). We permitted unrestricted uptake of sucrose as relative amounts of amino acid and Suc in maize peduncle is unclear. To ensure robustness of this approach with respect to variations of the user defined parameter α, we performed the simulations for values of $\alpha = 0.15$ and $\alpha = 0.3$. The simulations were performed to predict zein synthesis five different time points of grain filling. Nutrient composition and the structure of the metabolic network, which depends upon the user defined parameter α, at the five timepoints are vastly different. None of these simulations predicted synthesis of zeins, or starch (**Figure 5.3**).

We repeated the single level simulations by incorporating an additional constraint of a minimum amount of starch synthesis, with all other constraints and the objective remaining unchanged. The single level formulation predicted starch synthesis, due to the enforced constraint. However, zein synthesis did not occur unless excretion of specific amino acids were blocked. If excretion of all amino acids were blocked, the formulation could not find a feasible solution.

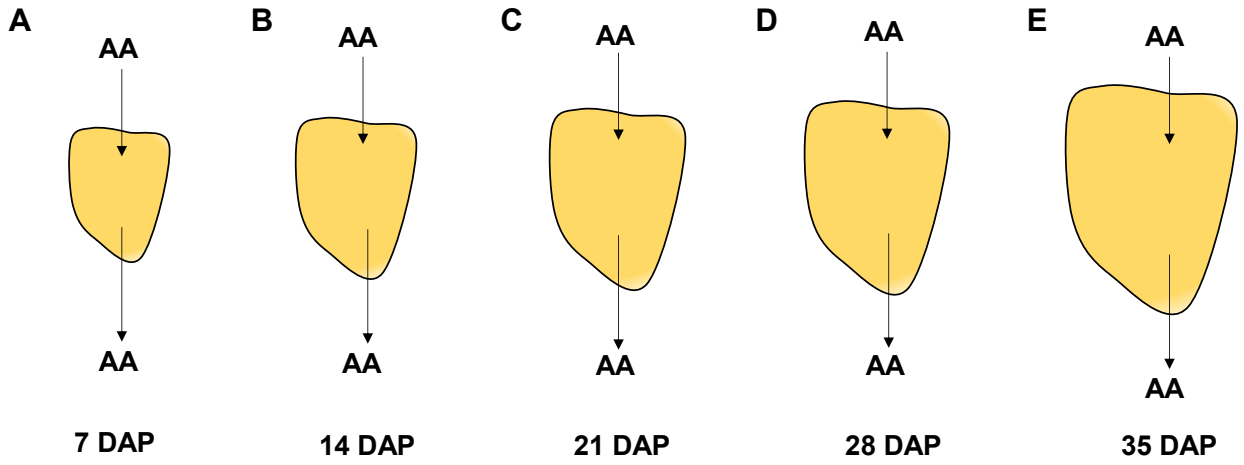


Figure 5.3. Predictions of endosperm metabolism with transcriptome matching as the sole objective. We investigated if a single level formulation employing gene expression is sufficient to accurately predict onset of specific zeins during grain filling. We constrained the formulation with an experimentally measured distribution of amino acids in vascular sap of maize peduncle (Arruda and Da Silva, 1979). We permitted unrestricted uptake of sucrose as relative amounts of amino acid and Suc in maize peduncle is unclear. We employed maximization of the transcriptome matching function as objective. We performed the simulations for two different values of the user defined parameter, $\alpha = 0.15$ and $\alpha = 0.3$. None of the simulations predicted accumulation of zeins. Arrows indicate uptake and secretion of metabolites by the model. A, prediction by single level formulation at 7 DAP. B, prediction by single level formulation at 14 DAP. C, prediction by single level formulation at 21 DAP. D, prediction by single level formulation at 28 DAP. E, prediction by single level formulation at 35 DAP.

5.4.2 Bi-level optimization with maximum nitrogen storage in zeins as the inner objective

The previous simulation clearly showed that single level formulation without *a priori* assumption of biological objective fails to capture cellular metabolism. Hence, we employed the bilevel framework with maximization of the transcriptome matching function as the outer objective and a carefully chosen objective as the inner objective function. The nutrient constraints remained identical to the single level formulation. To ensure robustness of this approach with respect to variations of the user defined parameter α, we performed the simulations for values of $\alpha = 0.15$ and $\alpha = 0.3$. The simulations were performed to predict zein synthesis five different time points of grain filling.

The first series of simulations with the bilevel framework employed the objective of maximum storage of N in zeins. With incorporation of the inner objective, the bilevel framework predicted zein synthesis at all time points (**Figure 5.4**). However, starch synthesis was not predicted at any stage. Additionally, the framework failed to predict synthesis of 16 kDa γ-zein, 19 kDa α-zein and 10 kDa δ-zein are not synthesized at any stage. The prediction of synthesis of 22 kDa α-zein 7 DAP is not supported by prior observations.

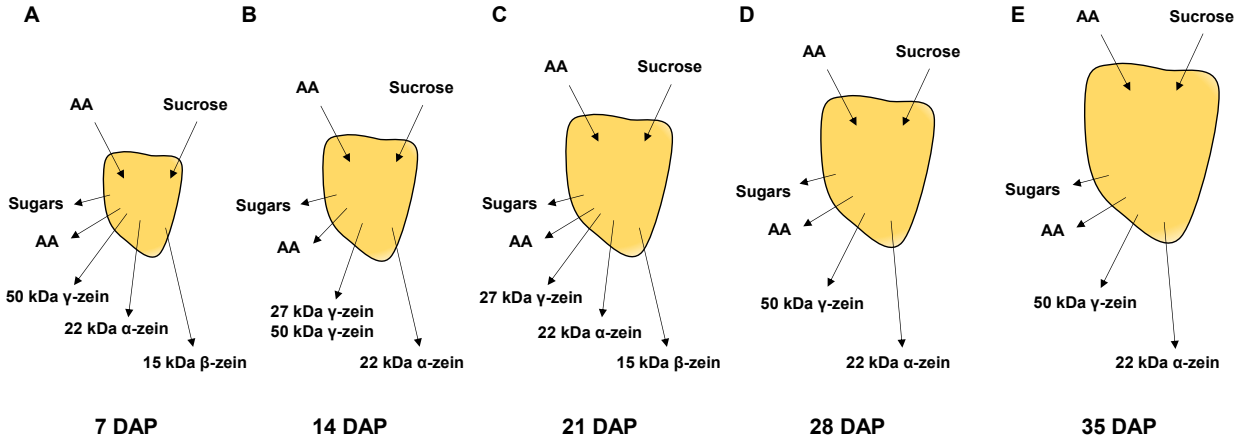


Figure 5.4. Predictions of endosperm metabolism with a bi-level optimization framework with maximum storage of N in zeins as the inner objective. We investigated if a bi-level optimization framework with maximum storage of nitrogen in zeins as the inner objective, and transcriptome matching as the outer objective, can accurately predict the onset of specific zeins during grain filling. We constrained the formulation with an experimentally measured distribution of amino acids in vascular sap of maize peduncle (Arruda and Da Silva, 1979). We permitted unrestricted uptake of sucrose as relative amounts of amino acid and Suc in maize peduncle is unclear. Additionally, we imposed storage of minimum of 25% of the total N uptake in zeins as a constraint. We performed the simulations with the user defined parameter $\alpha = 0.3$. The framework predicted synthesis of an array of zeins and carbohydrates at different time points. However, synthesis of starch, 16 kDa γ-zein, 19 kDa α-zein and 10 kDa δ-zein was not predicted at any time point. Arrows indicate uptake and secretion of metabolites by the model. A, prediction at 7 DAP. B, prediction at 14 DAP. C, prediction at 21 DAP. D, prediction at 28 DAP. E, prediction at 35 DAP.

5.4.3 Bi-level optimization with maximum starch synthesis as the inner objective

We employed the bilevel framework with maximization of the transcriptome matching function as the outer objective and maximum starch synthesis as the inner objective function. Additionally, we relaxed the value of the user chosen parameter to $\alpha = 0.15$. The rest of the constraints remained unchanged. With starch synthesis as the inner objective, the bilevel framework predicted concurrent synthesis of starch and zeins at all time points (**Figure 5.5**). Starch formation began at 7 DAP, peaked at 21 DAP and continued through 35 DAP. However, starch was not the dominant carbohydrate produced at any stage except 21 DAP. The predicted pattern of starch accumulation is

similar to trends observed in field observations. The synthesis of only γ-zein at the early stage (7 DAP) of grain filling is an accurate prediction. However, synthesis of 19 kDa α-zein and 10 kDa δ-zein was not predicted at any time point.

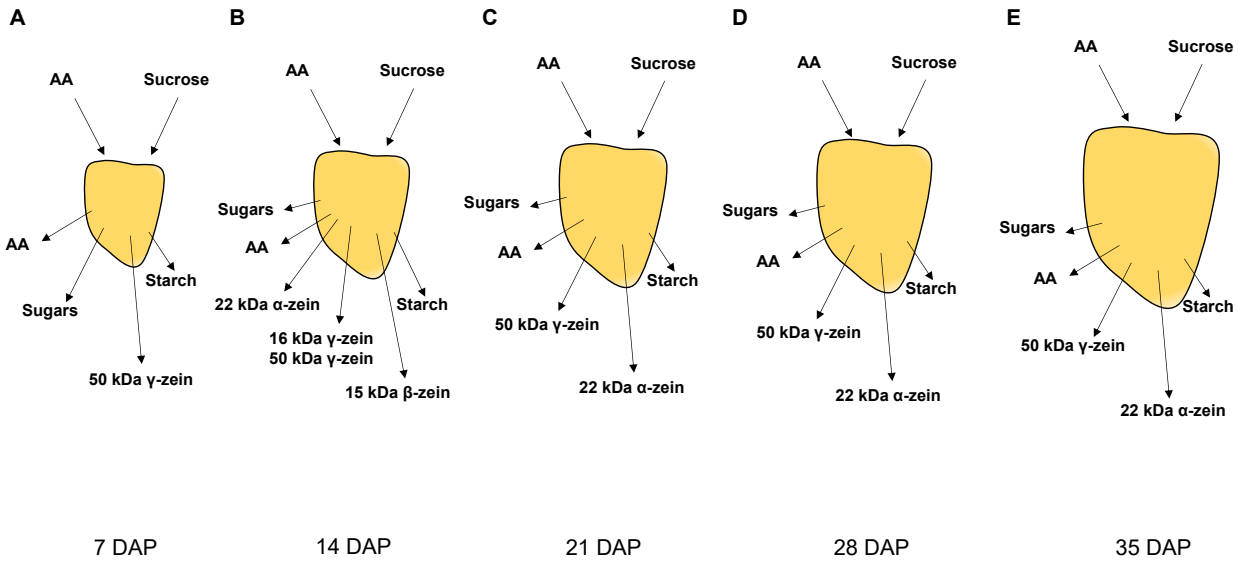


Figure 5.5. Predictions of endosperm metabolism with a bi-level optimization framework with maximum starch synthesis as the inner objective. We investigated if a bi-level optimization framework with maximum starch synthesis as inner objective, and transcriptome matching as the outer objective, can accurately predict the onset of specific zeins during grain filling. We constrained the formulation with an experimentally measured distribution of amino acids in vascular sap of maize peduncle (Arruda and Da Silva, 1979). We permitted unrestricted uptake of sucrose as relative amounts of amino acid and Suc in maize peduncle is unclear. Additionally, we imposed storage of minimum of 25% of the total N uptake in zeins as a constraint. We performed the simulations with the user defined parameter $\alpha = 0.15$. The framework predicted synthesis of starch and an array of zeins at different time points. However, synthesis of 19 kDa α-zein and 10 kDa δ-zein was not predicted at any time point. Arrows indicate uptake and secretion of metabolites by the model. A, prediction at 7 DAP. B, prediction at 14 DAP. C, prediction at 21 DAP. D, prediction at 28 DAP. E, prediction at 35 DAP.

5.4.4 Bi-level optimization with maximum Suc uptake as the inner objective

The last series of simulations with the bilevel framework employed maximization of the transcriptome matching function as the outer objective and maximum Suc uptake as the inner objective function. The value of the user chosen parameter was fixed at $\alpha = 0.15$. The rest of the constraints remained unchanged. With Suc uptake as the inner objective, the bilevel framework predicted concurrent synthesis of starch and zeins at all time points (**Figure 5.6**). The earliest timepoint featured predominately γ-zeins, but α-zeins dominated the later stages. Starch synthesis began at the time point of 14 DAP, peaked at 21 DAP and continued through 35 DAP. However, synthesis of β-zeins and the 10 kDa δ-zein was not predicted at any stage. However, starch was not the dominant carbohydrate produced at any stage except 21 DAP. The predicted pattern of starch

accumulation is similar to trends observed in field observations. The synthesis of γ-zein at the early stage (7 DAP) of grain filling is an accurate prediction.

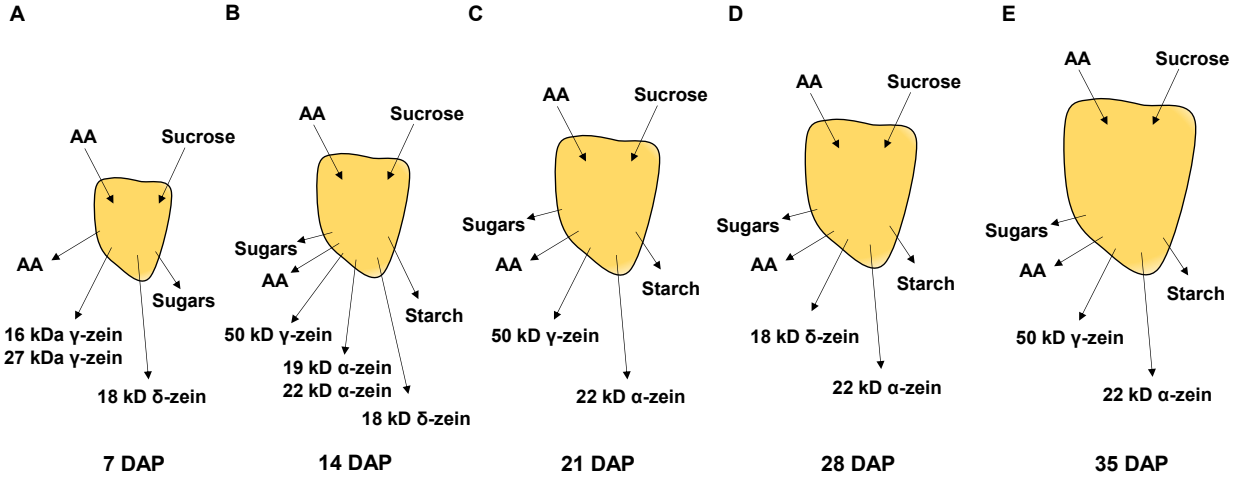


Figure 5.6. Predictions of endosperm metabolism with a bi-level optimization framework with maximum Suc uptake as the inner objective. We investigated if a bi-level optimization framework with maximum Suc uptake as inner objective, and transcriptome matching as the outer objective, can accurately predict the onset of specific zeins during grain filling. We constrained the formulation with an experimentally measured distribution of amino acids in vascular sap of maize peduncle (Arruda and Da Silva, 1979). We permitted unrestricted uptake of sucrose as relative amounts of amino acid and Suc in maize peduncle is unclear. Additionally, we imposed storage of minimum of 25% of the total N uptake in zeins as a constraint. We performed the simulations with the user defined parameter $\alpha = 0.15$. The framework predicted synthesis of starch and an array of zeins at different time points. However, synthesis of β-zeins and 10 kDa δ-zeins was not predicted at any time point. Arrows indicate uptake and secretion of metabolites by the model. A, prediction at 7 DAP. B, prediction at 14 DAP. C, prediction at 21 DAP. D, prediction at 28 DAP. E, prediction at 35 DAP.

5.5 Discussion

5.6.1 Improved predictive power of bilevel optimization

We have demonstrated that incorporation of a suitable inner objective function significantly improves the predictive power of gene-expression guided predictive modeling. With maximum starch synthesis as the inner objective, the bilevel formulation predicts the onset of starch synthesis and α-zein at 14 DAP respectively. The early accumulation of γ-zeins followed by accumulation of α-zeins is also supported by experimental observation. Out of the objectives tested, only the simulations with starch accumulation as the inner objective captured continual accumulation of γ-zeins. None of the bilevel formulations captured the synthesis of all the zeins. The relative quantities of β and δ-zeins in the endosperm is very small compared to α and γ-zeins.

The distribution of amino acids in α-zeins are remarkably similar (**Figure 5.1**). However, that is not the case in γ-zeins. The 50 kDA γ-zein contains significantly more Gln than any of the other zein proteins. As grain filling progresses, the amount of Gln in the nutrient decreases (**Figure 5.7**). Additionally, the transcript levels of genes associated with TCA cycles drop with progress in grain filling with a shift to glycolysis (Prioul Jean Louis et al., 2008). This results in excess Pyr generation through glycolysis, which leads to elevated synthesis of Leu rich α-zeins.

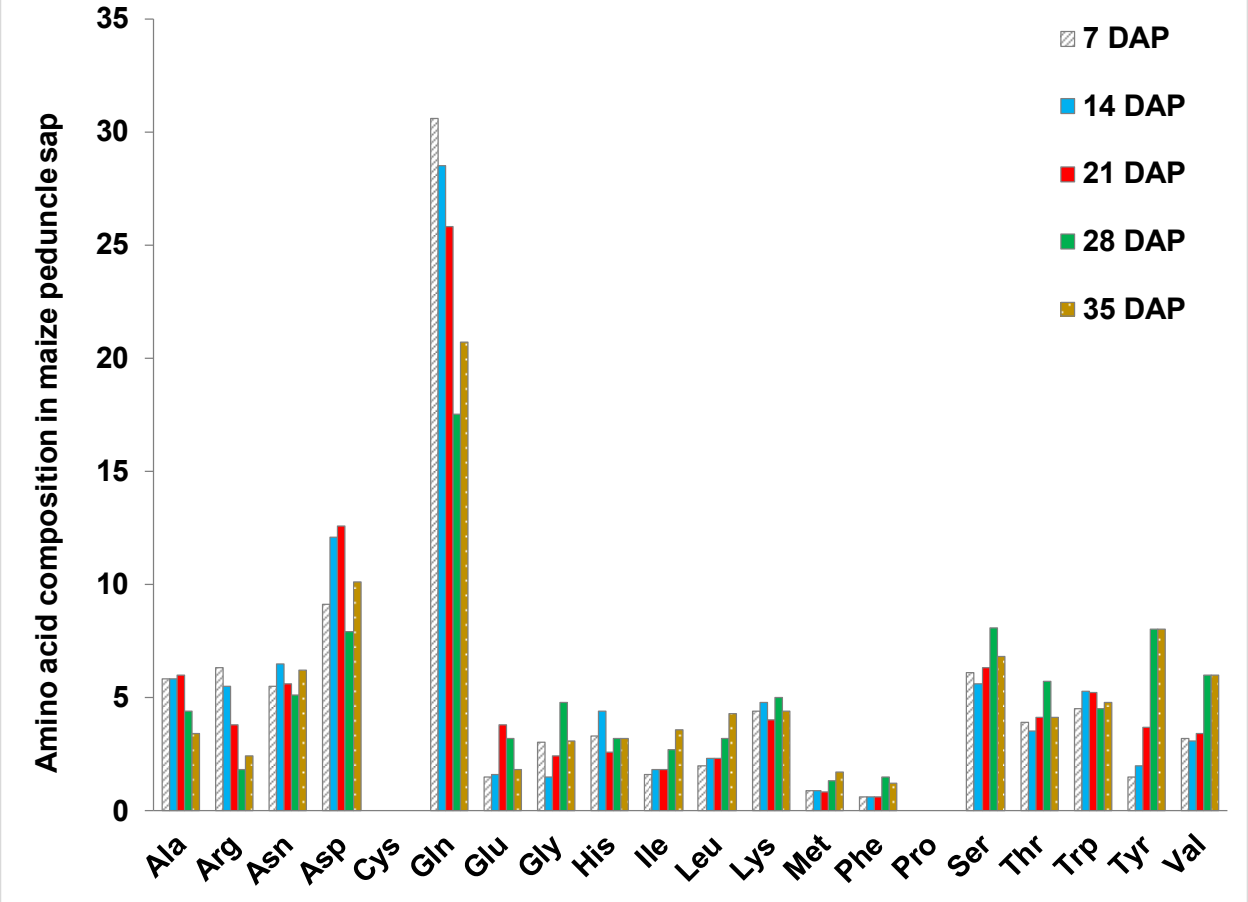


Figure 5.7. Distribution of all proteinogenic amino acids in at the five different time points. We extracted the data on experimentally measured amino acids composition of proteinogenic amino acids in the vascular sap of maize peduncle (Arruda and Da Silva, 1979). The plot shows the distributions at the five different timepoints of grain filling.

5.6.2 Iterative Approach Not Entirely Eliminated

Despite the improvement in predictive power the iterative approach to predictive modeling and analysis could not be eliminated. This is a result of having the inner objective as a linear combination of fluxes. As the fluxes included in the inner objective change, the conversion to the single level problem results in different sets of equations. When measurements as in case of maize endosperm is available, construction of the inner objective, either based on prior biochemical knowledge or observations of metabolite accumulation, is not very challenging. However, this remains a major bottleneck in investigations of tissues whose metabolism is relatively unknown.

5.6.3 Predictive modeling with transcriptome matching as the inner objective

We propose that changing the bilevel optimization framework to one in which the transcriptome matching function is the inner objective, and a linear objective function is the outer objective function will be more suitable for investigations in systems that are less well understood. However, the incorporation of binary (or integer) variables in the lower level objective function renders significant complexity as the duality theory cannot be applied to integer variables (Zeng and An). Only the last five years have seen the development of a handful of algorithms (Dempe and Kue, 2017; Yue et al., 2018; Zeng and An) that can be employed to solve this class of problems, often known as bilevel mixed integer programing problem. No commercial solver capable of directly solving bilevel mixed integer programing exists so far.

$$\max/\min\ z_{outer}$$

s.t.

$$\max\ z_{match} = \sum_{\forall j \in H} (d_j^{rev} + d_j^{for}) p_j - \sum_{\forall j \in L} (d_j^{rev} + d_j^{for})(1 - p_j)$$

s.t.

$$\begin{aligned}
&S_{ij} v_j = b_i && \forall j \\
&v_j \Delta g_j \le 0 && \forall j \notin \{\text{uptake, secrete}\} \\
&-v_j \le -lb_j d_j^{rev} && \forall j \notin \{\text{uptake, secrete}\} \\
&v_j \le ub_j d_j^{for} && \forall j \notin \{\text{uptake, secrete}\} \\
&-v_j \le -v_j^{lb} && \forall j \in \{\text{uptake}\} \\
&v_j \le v_j^{ub} && \forall j \in \{\text{uptake}\} \\
&-v_j \le 0 && \forall j \in \{\text{secrete}\} \\
&v_j \le v_j^{ub} && \forall j \in \{\text{secrete}\} \\
&d_j^{rev} + d_j^{for} \le 1,\ d_j^{rev}, d_j^{for} = \{0,1\}
\end{aligned}$$

The development of algorithms to solve bilevel mixed integer programing problems is much needed progress for not only systems biology but for operations research in general. However, the implementation of these algorithm is non-trivial (personal communication). Generally, these algorithms involve solving a series of relaxed subproblems to gradually close the bounds on the outer objective. The computational demand is intense, and these algorithms are unproven with large systems. We plan to apply bilevel mixed integer programing problem to predict metabolism with a plant GSM in future.

5.6 Conclusion

We developed a computational framework for integrating gene expression datasets with a genome-scale model of maize and used them to predict onset of zein protein accumulation during grain filling. The predictions indicate that the onset of accumulation of specific zeins is a response to decreasing O_2 in the hypoxic endosperm. Storage protein synthesis as a dual mechanism of N storage and alleviating the need for O_2 could potentially hold true for storage proteins in other grains as well. Additionally, modulating

the expression and deposition of the zein proteins might be a way of controlling kernel development. However, further studies will be necessary to validate these hypotheses.

Chapter 6 : Summary and Future Outlook

6.1 Summary

Predictive modeling with genome-scale stoichiometric metabolic networks is a powerful tool to develop nontrivial hypothesis on cellular metabolism and to explain insight on hitherto unexplained complex metabolic phenomena. In this dissertation, we applied predictive modeling techniques with the specific goals of: (1) assessing the metabolism of a cellulosic bacterium known to thrive in decaying organic matter in marine environment; (2) investigating internal seasonal recycling in a woody tree; and (3) investigating how storage tissue of a cereal fulfils its energetic need under hypoxia. The approaches employed in quest of these questions ranged from assessing the dual of a linear programing problem to construction of a multi-tissue genome-scale metabolic model and analysis based on multiple biologically relevant metrics. Additionally, we developed a bilevel optimization framework integrating gene expression data that can harness the advantage of generating transcriptome datasets from a wide array of plant tissues. The predictions resulting from these approaches are consistent with prior biochemical knowledge. The strategies demonstrated here were primarily driven by the challenge of acquiring enough data to constrain the models employed in the study. Here, predictive modeling for the first time has been employed to provide justifications on a well-known yet hitherto unexplained metabolic phenomena in trees. With the challenge of acquiring detailed experimental data, especially in plant systems, remaining unresolved for the foreseeable future, we anticipate a widespread application of the strategies adopted in our work in predictive modeling of plant metabolism.

6.2 Future Outlook

Cellular metabolism is a nonstationary process. Organisms have evolved to efficiently utilize available resources under changing environment to enhance their chances of survival. Hence, a steady-state model of cellular metabolism is only an approximation. However, estimation of metabolic fluxes under changing metabolism by nonstationary ^{13}C-MFA is computation intensive, which increases with network size and number of timepoints. This is compounded by the challenges associated with experimental aspects for transient labeling studies: controlled transient labeling of tissue of interest with a suitable isotope label, metabolite extraction and analysis of a very large number of samples at multiple timepoints to estimate isotope labeling patterns and concentration of metabolites, and the cost of such experiments. As a result, ^{13}C-MFA of seasonal metabolic events has not been realized yet.

Predictive modeling is significantly faster that flux estimation by nonstationary ^{13}C-MFA. However, the major challenge in predictive modeling of complex metabolic phenomena, especially those involving dynamic metabolic landscapes, is the lack of measurements that can be imposed as constraints in predictive modeling. The fidelity of

predictive modeling improves significantly with the quality of constraints enforced. Without such constraints, a predictive modeler must consider a wide range of scenarios and analyze a larger number of predictions, which slows down the analysis significantly. Additionally, predictions on complex metabolic phenomena often need to be benchmarked against qualitative data simply because nonstationary ^{13}C-MFA data of complex metabolic events is yet to be reported in the literature.

The bridge between predictive modeling and complete experimental verification be close by adopting an approach in which data collection is driven primarily by the objective of improving the quality of imposed constraints and verification of the predictions. We have demonstrated even in the absence of detailed data on nutrient uptake, meaningful insight can be obtained by careful formulation of the problems under investigation. However, it is desirable to validate the predictions against experimental measurements.

The strategies and methods demonstrated here can be employed in other organisms. The challenge however is in translation of the insights gained from predictive modeling to identifying specific targets for genetic manipulation to enhance a desired trait in an organism of interest. We anticipate that this work would expand the scope of application of predictive modeling to investigations of hitherto unexplained metabolic scenarios and help develop potential strategies for genetic modifications.

References

Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput Biol **8**: e1002518

Aharoni A, Galili G (2011) Metabolic engineering of the plant primary–secondary metabolism interface. Curr Opin Biotechnol **22**: 239–244

Åkesson M, Förster J, Nielsen J (2004) Integration of gene expression data into genome-scale metabolic models. Metab Eng **6**: 285–293

Andrykovitch G, Marx I (1988) Isolation of a new polysaccharide-digesting bacterium from a salt marsh. Appl Environ Microbiol **54**: 1061–1062

Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab Eng **9**: 68–86

Antonio C, Päpke C, Rocha M, Diab H, Limami AM, Obata T, Fernie AR, van Dongen JT (2015) Regulation of primary metabolism in response to low oxygen availability as revealed by carbon and nitrogen isotope redistribution. Plant Physiol **170**: 43–56

Arruda P, Da Silva WJ (1979) Amino acid composition of vascular sap of maize ear peduncle. Phytochemistry **18**: 409–410

Babst BA, Coleman GD (2018) Seasonal nitrogen cycling in temperate trees: transport and regulatory mechanisms are key missing links. Plant Sci **270**: 268–277

Baghalian K, Hajirezaei M-R, Schreiber F (2014) Plant metabolic modeling: achieving new insight into metabolism and metabolic engineering. Plant Cell **26**: 3847–3866

Bauer JE, Cai W-J, Raymond PA, Bianchi TS, Hopkinson CS, Regnier PAG (2013) The changing carbon cycle of the coastal ocean. Nature **504**: 61–70

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc **2**: 727–738

Becker SA, Palsson BØ (2008) Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol **4**: e1000082

Boehlein SK, Shaw JR, Stewart JD, Hannah LC (2010) Studies of the kinetic mechanism of maize endosperm ADP-glucose pyrophosphorylase uncovered complex regulatory properties. Plant Physiol **152**: 1056–1064

Bogart E, Myers CR (2016) Multiscale metabolic modeling of C4 plants: connecting nonlinear genome-scale models to leaf-scale metabolism in developing maize leaves. PLoS One **11**: e0151722

Bordbar A, Monk JM, King ZA, Palsson BØ (2014) Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet **15**: 107–120

Borisjuk L, Rolletschek H (2009) The oxygen status of the developing seed. New Phytol **182**: 17–30

Bowsher CG, Scrase-Field EFAL, Esposito S, Emes MJ, Tetlow IJ (2007) Characterization of ADP-glucose transport across the cereal endosperm amyloplast envelope. J Exp Bot **58**: 1321–1332

Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A **107**: 17845–17850

Chastain CJ, Failing CJ, Manandhar L, Zimmerman MA, Lakner MM, Nguyen THT (2011) Functional evolution of C_4 pyruvate, orthophosphate dikinase. J Exp Bot **62**: 3083–3091

Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol **166**: 252–264

Chen X, Schreiber K, Appel J, Makowka A, Fähnrich B, Roettger M, Hajirezaei MR, Sönnichsen FD, Schönheit P, Martin WF, et al (2016) The Entner–Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. Proc Natl Acad Sci U S A **113**: 5441–5446

Cheung CYM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ (2014) A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C_3 and crassulacean acid metabolism leaves. Plant Physiol **165**: 917–929

Cheung CYM, Williams TCR, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. Plant J **75**: 1050–1061

Churkina G (2016) The role of urbanization in the global carbon cycle. Front Ecol Evol **3**. doi:10.3389/fevo.2015.00144

Coleman GD, Chen THH, Fuchigami LH (1992) Complementary DNA cloning of poplar bark storage protein and control of Its expression by photoperiod. Plant Physiol **98**: 687–693

Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng T-Y, Moody DB, Murray M, Galagan JE (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput Biol **5**: e1000489

Colón AM, Sengupta N, Rhodes D, Dudareva N, Morgan J (2010) A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of *Petunia hybrida*. Plant J **62**: 64–76

Cooke JEK, Weih M (2005) Nitrogen storage and seasonal nitrogen cycling in *Populus*: bridging molecular physiology and ecophysiology. New Phytol **167**: 19–30

Dal'Molin CG de O, Nielsen LK (2018) Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses. Curr Opin Biotechnol **49**: 42–48

Dal'Molin CG de O, Nielsen LK (2012) Plant genome-scale metabolic reconstruction and modelling. Curr Opin Biotechnol **24**: 1–7

Dal'Molin CG de O, Orellana C, Gebbie L, Steen J, Hodson MP, Chrysanthopoulos P, Plan MR, McQualter R, Palfreyman RW, Nielsen LK (2016) Metabolic reconstruction of *Setaria italica*: a systems biology approach for integrating tissue-specific omics and pathway analysis of bioenergy grasses. Front Plant Sci **7**. doi:10.3389/fpls.2016.01138

Dal'Molin CG de O, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK (2010a) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol **152**: 579–589

Dal'Molin CG de O, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK (2010b) C4GEM, a genome-scale metabolic model to study C_4 plant metabolism. Plant Physiol **154**: 1871–1885

Dal'Molin CG de O, Quek L-E, Saa PA, Nielsen LK (2015) A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. Front Plant Sci. doi:10.3389/fpls.2015.00004

Dal'Molin DO, Gomes C, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK (2010c) AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. Plant Physiol **152**: 579–589

Dellero Y, Jossier M, Schmitz J, Maurino VG, Hodges M (2016) Photorespiratory glycolate–glyoxylate metabolism. J Exp Bot **67**: 3041–3052

Dempe S, Kue FM (2017) Solving discrete linear bilevel optimization problems using the optimal value reformulation. J Glob Optim **68**: 255–277

Denyer K, Dunlap F, Thorbjørnsen T, Keeling P, Smith AM (1996) The major form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. Plant Physiol **112**: 779–785

Druart N, Johansson A, Baba K, Schrader J, Sjödin A, Bhalerao RR, Resman L, Trygg J, Moritz T, Bhalerao RP (2007) Environmental and hormonal regulation of the activity–dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. Plant J **50**: 557–573

Dyksma S, Bischof K, Fuchs BM, Hoffmann K, Meier D, Meyerdierks A, Pjevac P, Probandt D, Richter M, Stepanauskas R, et al (2016) Ubiquitous *Gammaproteobacteria* dominate dark carbon fixation in coastal sediments. ISME J **10**: 1939–1953

Ebrahim A, Lerman JA, Palsson BØ, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol **7**: 74

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300**: 1005–1016

Emes MJ, Bowsher CG, Hedley C, Burrell MM, Scrase-Field ESF, Tetlow IJ (2003) Starch synthesis and carbon partitioning in developing endosperm. J Exp Bot **54**: 569–575

Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Gall SL, et al (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol Cell Proteomics **9**: 1063–1084

Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal Biochem **325**: 308–16

Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci U S A **110**: 10039–10044

Ganesh Sriram, D. Bruce Fulton, Vidya V. Iyer, Joan Marie Peterson, Ruilian Zhou, Mark E. Westgate, Martin H. Spalding, Jacqueline V. Shanks (2004) Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing. Plant Physiol **136**: 3043–3057

Gayral M, Elmorjani K, Dalgalarrondo M, Balzergue SM, Pateyron S, Morel M-H, Brunet S, Linossier L, Delluc C, Bakan B, et al (2017) Responses to hypoxia and endoplasmic reticulum stress discriminate the development of vitreous and

floury endosperms of conventional maize (*Zea mays*) inbred lines. Front Plant Sci **8**. doi:10.3389/fpls.2017.00557

Gianazza E, Viglienghi V, Righetti PG, Salamini F, Soave C (1977) Amino acid composition of zein molecular components. Phytochemistry **16**: 315–317

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet **21**: 227–233

Goldberg AP, Szigeti B, Chew YH, Sekar JA, Roth YD, Karr JR (2018) Emerging whole-cell modeling principles and methods. Curr Opin Biotechnol **51**: 97–102

Grafahrend-Belau E, Junker A, Eschenröder A, Müller J, Schreiber F, Junker BH (2013) Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. Plant Physiol **163**: 637–647

Grafahrend-Belau E, Schreiber F, Koschützki D, Junker BH (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant Physiol **149**: 585–598

Guo X, Yuan L, Chen H, Sato SJ, Clemente TE, Holding DR (2013) Nonredundant function of zeins and their correct stoichiometric ratio drive protein body formation in maize endosperm. Plant Physiol **162**: 1359–1369

Heazlewood JL, Millar AH (2005) AMPDB: the Arabidopsis mitochondrial protein database. Nucl Acids Res **33**: D605–D610

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucl Acids Res **35**: W585–W587

Huang B, Hennen-Bierwagen TA, Myers AM (2014) Functions of multiple genes encoding ADP-glucose pyrophosphorylase subunits in maize endosperm, embryo, and leaf. Plant Physiol **164**: 596–611

Hügler M, Sievert SM (2011) Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Ann Rev Mar Sci **3**: 261–289

James MG, Denyer K, Myers AM (2003) Starch synthesis in the cereal endosperm. Curr Opin Plant Biol **6**: 215–222

Jazmin LJ, O'Grady JP, Ma F, Allen DK, Morgan JA, Young JD (2014) Isotopically nonstationary MFA (INST-MFA) of autotrophic metabolism. Plant Metabolic Flux Analysis: Methods and Protocols. Eds. Dieuaide-Noubhani M and Alonso AP. Humana Press, Totowa, NJ: 181–210

Jensen PA, Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics **27**: 541–547

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L (2000) The large-scale organization of metabolic networks. Nature **407**: 651–654

Karp PD, Riley M, Paley SM, Pellegrini-Toole A (2002) The MetaCyc database. Nucl Acids Res **30**: 59–61

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr. B, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. Cell **150**: 389–401

Kathilankal JC, Mozdzer TJ, Fuentes JD, D'Odorico P, McGlathery KJ, Zieman JC (2008) Tidal influences on carbon assimilation by a salt marsh. Environ Res Lett **3**: 044010

Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol **14**: 491–496

Kesselmeier J, Staudt M (1999) Biogenic volatile organic compounds (VOC): an overview on emission, physiology and ecology. J Atmos Chem **33**: 23–88

Kim MK, Lun DS (2014) Methods for integration of transcriptomic data in genome-scale metabolic models. Comput Struct Biotechnol J **11**: 59–65

Kirwan ML, Mudd SM (2012) Response of salt-marsh carbon accumulation to climate change. Nature **489**: 550–553

Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis? Trends Biotechnol **21**: 64–69

Klingner A, Bartsch A, Dogs M, Wagner-Döbler I, Jahn D, Simon M, Brinkhoff T, Becker J, Wittmann C (2015) Large-scale ^{13}C flux profiling reveals conservation of the Entner-Doudoroff pathway as a glycolytic strategy among marine bacteria that use glucose. Appl Environ Microbiol **81**: 2408–2422

Kruger NJ, Masakapalli SK, Ratcliffe RG (2011) Strategies for investigating the plant metabolic network with steady-state metabolic flux analysis: lessons from an Arabidopsis cell culture and other systems. J of Exp Bot **63**: 2309–2323

Kumar V, Baweja M, Singh PK, Shukla P (2016) Recent developments in systems biology and metabolic engineering of plant–microbe interactions. Front Plant Sci **7**. doi:10.3389/fpls.2016.01421

Kumar VS, Dasika MS, Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics **8**: 212

Lakshmanan M, Cheung CYM, Mohanty B, Lee D-Y (2016) Modeling rice metabolism: from elucidating environmental effects on cellular phenotype to guiding crop improvement. Front Plant Sci **7**. doi:10.3389/fpls.2016.01795

Lakshmanan M, Koh G, Chung BKS, Lee D-Y (2014) Software applications for flux balance analysis. Brief Bioinform **15**: 108–122

Lappe RR, Baier JW, Boehlein SK, Huffman R, Lin Q, Wattebled F, Settles AM, Hannah LC, Borisjuk L, Rolletschek H, Stewart JD, Scott MP, Hennen-Bierwagen TA, Myers AM (2018) Functions of maize genes encoding pyruvate phosphate dikinase in developing endosperm. Proc Natl Acad Sci U S A **115**: E24–E33

Lau W, Fischbach MA, Osbourn A, Sattely ES (2014) Key applications of plant metabolic engineering. PLoS Biol **12**: e1001879

Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, Mendes P, Swainston N (2012) Improving metabolic flux predictions using absolute gene expression data. BMC Syst Biol **6**: 73

Leithold EL, Blair NE, Wegmann KW (2016) Source-to-sink sedimentary systems and global carbon burial: a river runs through it. Earth Sci Rev **153**: 30–42

Levering J, Broddrick J, Dupont CL, Peers G, Beeri K, Mayers J, Gallina AA, Allen AE, Palsson BØ, Zengler K (2016) Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PLoS One **11**: e0155038

Lloyd CJ, Ebrahim A, Yang L, King ZA, Catoiu E, O'Brien EJ, Liu JK, Palsson BØ (2018) COBRAme: A computational framework for genome-scale models of metabolism and gene expression. PLoS Comput Biol **14**: e1006302

Logan BA, Monson RK, Potosnak MJ (2000) Biochemistry and physiology of foliar isoprene production. Trends Plant Sci **5**: 477–481

Loreto F, Pinelli P, Brancaleoni E, Ciccioli P (2004) ^{13}C labeling reveals chloroplastic and extrachloroplastic pools of dimethylallyl pyrophosphate and their contribution to isoprene formation. Plant Physiol **135**: 1903–1907

Machado D, Herrgård M (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Computat Biol **10**: e1003580

Macklin DN, Ruggero NA, Covert MW (2014) The future of whole-cell modeling. Curr Opin Biotechnol **28**: 111–115

Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng **5**: 264–276

Maranas, Zomorrodi AR (2016) Optimization Methods in Metabolic Networks. Wiley-Blackwell

Méchin V, Thévenot C, Le Guilloux M, Prioul J-L, Damerval C (2007a) Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. Plant Physiol **143**: 1203–1219

Millard P, Grelet G (2010) Nitrogen storage and remobilization by trees: ecophysiological relevance in a changing world. Tree Physiol **30**: 1083–1095

Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proc Natl Acad Sci U S A **109**: 339–344

Misra S, Oaks A (1985) Glutamine metabolism in corn kernels cultured *in vitro*. Plant Physiol **77**: 520–523

Monaco MK, Sen TZ, Dharmawardhana PD, Ren L, Schaeffer M, Naithani S, Amarasinghe V, Thomason J, Harper L, Gardiner J, et al (2013) Maize metabolic network construction and transcriptome analysis **6**. Plant Genome. doi:10.3835/plantgenome2012.09.0025

Morrison WR (1988) Lipids in cereal starches: a review. J Cereal Sci **8**: 1–15

Moxley JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper H, Wheeler RT, Tong L, Hinnebusch AG, Ideker T, Nielsen J, Stephanopoulos G (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc Natl Acad Sci U S A **106**: 6477–6482

Muhitch MJ (1988) Acetolactate synthase activity in developing maize (*Zea mays* L.) kernels. Plant Physiol **86**: 23–27

Nargund S, Misra A, Zhang X, Coleman GD, Sriram G (2014) Flux and reflux: metabolite reflux in plant suspension cells and its implications for isotope-assisted metabolic flux analysis. Mol Biosyst **10**: 1496

O'Connor SE (2015) Engineering of secondary metabolism. Annu Rev Genet **49**: 71–94

Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE (2017) A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Syst **4**: 318-329

Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotech **28**: 245–248

Pilalis E, Chatziioannou A, Thomasset B, Kolisis F (2011) An in silico compartmentalized metabolic model of *Brassica napus* enables the systemic study of regulatory aspects of plant central metabolism. Biotechnol Bioeng **108**: 1673–1682

Poolman MG, Kundu S, Shaw R, Fell DA (2013) Responses to light intensity in a genome-scale model of rice metabolism. Plant Physiol **162**: 1060–1072

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of Arabidopsis thaliana and some of its properties. Plant Physiol **151**: 1570–1581

Prioul Jean Louis, Méchin Valérie, Lessard Philippe, Thévenot Claudine, Grimmer Mike, Chateau-Joubert Sophie, Coates Steve, Hartings Hans, Kloiber-Maitz Monica, Murigneux Alain, et al (2008) A joint transcriptomic, proteomic and metabolic analysis of maize endosperm development and starch filling. Plant Biotechnol J **6**: 855–869

Quin MB, Schmidt-Dannert C (2014) Designer microbes for biosynthesis. Curr Opin Biotechnol **29**: 55–61

Rasulov B, Copolovici L, Laisk A, Niinemets Ü (2009) Postillumination isoprene emission: in vivo measurements of dimethylallyldiphosphate pool size and isoprene synthase kinetics in aspen leaves. Plant Physiol **149**: 1609–1618

Reed JL (2012) Shrinking the metabolic solution space using experimental datasets. PLoS Comput Biol **8**: e1002662

Reed JL, Vo TD, Schilling CH, Palsson BØ (2003) An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). Genome Biol **4**: R54

Rennenberg H, Wildhagen H, Ehlting B (2010) Nitrogen nutrition of poplar trees. Plant Biol **12**: 275–291

Reumann S, Ma C, Lemke S, Babujee L (2004) AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. Plant Physiol **136**: 2587–2608

Riebesell U, Schulz KG, Bellerby RGJ, Botros M, Fritsche P, Meyerhöfer M, Neill C, Nondal G, Oschlies A, Wohlers J, Zo¨llner E (2007) Enhanced biological carbon consumption in a high CO_2 ocean. Nature **450**: 545–548

Rios-Estepa R, Lange BM (2007) Experimental and mathematical approaches to modeling plant metabolic networks. Phytochemistry **68**: 2351–2374

Rolletschek H, Koch K, Wobus U, Borisjuk L (2005) Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. Plant J **42**: 69–83

Rolletschek H, Melkus G, Grafahrend-Belau E, Fuchs J, Heinzel N, Schreiber F, Jakob PM, Borisjuk L (2011) Combined noninvasive imaging and modeling approaches reveal metabolic compartmentation in the barley endosperm. Plant Cell **23**: 3041–3054

Sabelli PA, Larkins BA (2009) The development of endosperm in grasses. Plant Physiol **149**: 14–26

Saha R, Chowdhury A, Maranas CD (2014) Recent advances in the reconstruction of metabolic models and integration of omics data. Curr Opin Biotechnol **29**: 39–45

Saha R, Suthers PF, Maranas CD (2011) Zea mays *i*RS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS One **6**: e21784

Sannigrahi P, Ragauskas AJ, Tuskan GA (2010) Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuel Bioprod Biorefin **4**: 209–226

Sauter JJ, van Cleve B (1992) Seasonal variation of amino acids in the xylem sap of "Populus x canadensis" and its relation to protein body mobilization. Trees **7**: 26–32

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc **6**: 1290–1307

Scheunemann M, Brady SM, Nikoloski Z (2018) Integration of large-scale data for extraction of integrated Arabidopsis root cell-type specific models. Sci Rep **8**: 7919

Schilmiller AL, Pichersky E, Last RL (2012) Taming the hydra of specialized metabolism: how systems biology and comparative approaches are revolutionizing plant biochemistry. Curr Opin Plant Biol **15**: 338–344

Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR (2013) GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics **29**: 2900–2908

Schwender J, König C, Klapperstück M, Heinzel N, Munz E, Hebbelmann I, Hay JO, Denolf P, De Bodt S, Redestig H, Caestecker E, Jakob PM, Borisjuk L, Rolletschek H (2014) Transcript abundance on its own cannot be used to infer fluxes in central metabolism. Front Plant Sci **5**. doi:10.3389/fpls.2014.00668

Seaver SMD, Bradbury LMT, Frelin O, Zarecki R, Ruppin E, Hanson AD, Henry CS (2015) Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. Front Plant Sci **6**. doi:10.3389/fpls.2015.00142

Seaver SMD, Henry CS, Hanson AD (2012) Frontiers in metabolic reconstruction and modeling of plant genomes. J Exp Bot **63**: 2247–2258

Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: why and how. Ann Bot **101**: 5–18

Sharkey TD, Yeh S (2001) Isoprene emission from plants. Annu Rev Plant Physiol Plant Mol Biol **52**: 407–436

Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot **53**: 947–958

Shi H, Schwender J (2016a) Mathematical models of plant metabolism. Curr Opin Biotechnol **37**: 143–152

Shi H, Schwender J (2016b) Mathematical models of plant metabolism. Curr Opin Biotechnol **37**: 143–152

Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotech **26**: 1003–1010

Shukla R, Cheryan M (2001) Zein: the industrial protein from corn. Industrial Crops and Products **13**: 171–192

Simons M, Misra A, Sriram G (2014a) Genome-scale models of plant metabolism. Plant Metabolism: Methods and Protocols. Ed. Sriram G. Humana Press, Totowa, NJ: 213–230

Simons M, Saha R, Amiour N, Kumar A, Guillard L, Clément G, Miquel M, Li Z, Mouille G, Lea PJ, et al (2014b) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol **166**: 1659–1674

Smith SV, Hollibaugh JT (1993) Coastal metabolism and the oceanic organic carbon balance. Rev Geophys **31**: 75–89

Sodek L, da Silva WJ (1977) Glutamate synthase: a possible role in nitrogen metabolism of the developing maize endosperm. Plant Physiol **60**: 602–605

Song H-S, Reifman J, Wallqvist A (2014) Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. PLoS One **9**: e112524

Spielbauer G, Li L, Römisch-Margl L, Do PT, Fouquet R, Fernie AR, Eisenreich W, Gierl A, Settles AM (2013) Chloroplast-localized 6-phosphogluconate dehydrogenase is critical for maize endosperm starch accumulation. J Exp Bot **64**: 2231–2242

Sriram G (2004) Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping. Metab Eng **6**: 116–132

Suárez MF, Avila C, Gallardo F, Cantón FR, García-Gutiérrez A, Claros MG, Cánovas FM (2002) Molecular and enzymatic analysis of ammonium assimilation in woody plants. J Exp Bot **53**: 891–904

Sundareshwar P, Morris J, Koepfler E, Fornwalt B (2003) Phosphorus limitation of coastal ecosystem processes. Science **299**: 563–5

Swainston N, Smallbone K, Mendes P, Kell D, Paton N (2011) The SuBliMinaL toolbox: automating steps in the reconstruction of metabolic networks. J Integr Bioinform **8**. doi:10.1515/jib-2011-186

Szecowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M, Fernie AR, Arrivault S (2013) Metabolic fluxes in an illuminated Arabidopsis rosette. Plant Cell **25**: 694-714

Tatsis EC, O'Connor SE (2016) New developments in engineering plant metabolic pathways. Curr Opin Biotechnoly **42**: 126–132

Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? Ann Bot **90**: 681–689

Thompson GA, Larkins BA (1994) Characterization of zein genes and their regulation in maize endosperm. The Maize Handbook. Springer, New York, NY, pp 639–647

Töpfer N, Caldana C, Grimbs S, Willmitzer L, Fernie AR, Nikoloski Z (2013) Integration of genome-scale modeling and transcript profiling reveals metabolic pathways underlying light and temperature acclimation in Arabidopsis. Plant Cell **25**: 1197–1211

Töpfer N, Jozefczuk S, Nikoloski Z (2012) Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. BMC Syst Biol **6**: 148

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596–1604

van Berlo R, de Ridder D, Daran J, Daran-Lapujade P, Teusink B, Reinders M (2011) Predicting metabolic fluxes using gene expression differences as constraints. IEEE/ACM Trans Comput Biol Bioinform **8**: 206–216

Vance CP, Uhde-Stone C, Allan DL (2003) Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. New Phytol **157**: 423–447

Vranová E, Coman D, Gruissem W (2012) Structure and dynamics of the isoprenoid pathway network. Mol Plant **5**: 318–333

Walley JW, Shen Z, Sartor R, Wu KJ, Osborn J, Smith LG, Briggs SP (2013) Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc Natl Acad Sci U S A **110**: E4808–E4817

Walsh JR, Schaeffer ML, Zhang P, Rhee SY, Dickerson JA, Sen TZ (2016) The quality of metabolic pathway resources depends on initial enzymatic function assignments: a case for maize. BMC Syst Biol **10**: 129

Wang G, Zhang J, Wang G, Fan X, Sun X, Qin H, Xu N, Zhong M, Qiao Z, Tang Y, Song R (2014) Proline responding plays a critical role in regulating general protein synthesis and the cell cycle in maize. Plant Cell **26**: 2582–2600

Wang L, Dash S, Ng CY, Maranas CD (2017) A review of computational tools for design and reconstruction of metabolic pathways. Synth Syst Biotechnol **2**: 243–252

Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L (2002) Gramene: a resource for comparative grass genomics. Nucl Acids Res **30**: 103–105

Weiner RM, Taylor LE, Henrissat B, Hauser L, Land M, Coutinho PM, Rancurel C, Saunders EH, Longmire AG, Zhang H, et al (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40T. PLoS Genet **4**. doi:10.1371/journal.pgen.1000087

Wiberley AE, Donohue AR, Meier ME, Westphal MM, Sharkey TD (2008) Regulation of isoprene emission in *Populus trichocarpa* leaves subjected to changing growth temperature. Plant Cell Environ **31**: 258–267

Wiechert W (2001) ^{13}C metabolic flux analysis. Metab Eng **3**: 195–206

Wiechert W, Möllney M, Isermann N, Wurzel M, de Graaf AA (1999) Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol Bioeng **66**: 69–85

Wu Y, Holding DR, Messing J (2010) γ-Zeins are essential for endosperm modification in quality protein maize. Proc Natl Acad Sci U S A **107**: 12810–12815

Yan D, Duermeyer L, Leoveanu C, Nambara E (2014) The Functions of the Endosperm During Seed Germination. Plant Cell Physiol **55**: 1521–1533

Young JD, Shastri AA, Stephanopoulos G, Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis. Metab Eng **13**: 656–665

Yuan H, Cheung CYM, Hilbers PAJ, Riel V, W NA (2016a) Flux balance analysis of plant metabolism: the effect of biomass composition and model structure on model predictions. Front Plant Sci **7**. doi:10.3389/fpls.2016.00537

Yuan H, Cheung CYM, Poolman MG, Hilbers PAJ, Riel NAW van (2016b) A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. Plant J **85**: 289–304

Yue D, Gao J, Zeng B, You F (2018) A projection-based reformulation and decomposition algorithm for global optimization of a class of mixed integer bilevel linear programs. J Glob Optim: 1–31

Zamboni N, Fendt S-M, Ruhl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protocols **4**: 878–892

Zeng B, An Y (2014) Solving bilevel mixed integer program by reformulations and decomposition. Optimization Online

Zhang H, Hutcheson SW (2011) Complex expression of the cellulolytic transcriptome of *Saccharophagus degradans*. Appl Environ Microbiol **77**: 5591–5596

Zhang P, Dreher K, Karthikeyan A, Chi A, Pujar A, Caspi R, Karp P, Kirkup V, Latendresse M, Lee C, et al (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. Plant Physiol **153**: 1479–1491

Zhang X, Misra A, Nargund S, Coleman GD, Sriram G (2018) Concurrent isotope-assisted metabolic flux analysis and transcriptome profiling reveal responses of poplar cells to altered nitrogen and carbon supply. Plant J **93**: 472–488

Zheng Y and Sriram G (2014) Steady-state and instationary modeling of proteinogenic and free amino acid isotopomers for flux quantification. Plant Metabolic Flux Analysis: Methods and Protocols. Eds. Dieuaide-Noubhani M and Alonso AP. Humana Press, Totowa, NJ: 155–179

Zhang Z, Yang J, Wu Y (2015) Transcriptional regulation of zein gene expression in maize through the additive and synergistic action of opaque2, prolamine-box binding factor, and O2 heterodimerizing proteins. Plant Cell **27**: 1162–1172