

**UNRAVELING THE REGULATION OF MINT MONOTERPENE  
BIOSYNTHESIS: DEVELOPMENT AND EXPERIMENTAL TESTING OF  
KINETIC MATHEMATICAL MODELS**

By

RIGOBERTO RIOS-ESTEPA

A dissertation submitted in partial fulfillment of  
the requirements for the degree of


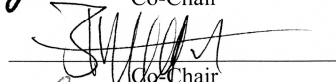
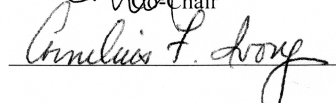

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY  
The Gene and Linda Voiland School of Chemical Engineering and Bioengineering

DECEMBER 2008

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of RIGOBERTO RIOS-ESTEPA find it satisfactory and recommend that it be accepted.

  
Co-Chair  
  
Co-Chair  
  


## ACKNOWLEDGEMENTS

First of all I would like to express my gratitude to my supervisors Dr. James Lee and Dr. Mark Lange for their guidance, continuous support and encouragement. My sincere and special thanks to Dr Lange, an outstanding scientist who I admire, for having given me the opportunity to perform all my experimental work in his lab, and for the chance of going into the challenging area of Systems Biology. I would also like to thank the other members of my graduate committee, Prof. Neil Ivory and Prof. David Kramer, for help and advice in many aspects of my program.

I owe many thanks to many current and past member of the Lange Lab for all kinds of help: Mr. Feldman, Siau Sie Voo, Cindy Meyrat, Dan Cuthbertson, special thanks to Dr. Glenn Turner and Iris Lange, for patiently teaching me and helping me with part of my experimental work.

I certainly owe a huge debt of gratitude to Dr. Rod Croteau and his lab members for providing us with the pulegone reductase clone and the transgenic plant lines. Special thanks to Drs. Ed Davis, Ray Ketchum and Kerry Ringer for very helpful comments and discussions. My thanks also go to Dr. Jeff Cruz in the Kramer Lab.

I am grateful to the Department of Chemical Engineering and its departmental staff: Jo Ann McCabe, Diana Thornton, and Senja Estes. Many thanks to the IBC staff, Helen Miller, Paula Gibbs, Beth Miller and Mary Kittelson. I am also grateful for the tremendous and invaluable support from Julia Gothard and Craig Whitney in the greenhouse facilities.

Finally, my deepest thanks go to my wife, Dalis, for her love, encouragement and companionship.

**UNRAVELING THE REGULATION OF MINT MONOTERPENE  
BIOSYNTHESIS: DEVELOPMENT AND EXPERIMENTAL TESTING OF  
KINETIC MATHEMATICAL MODELS**

Abstract

by Rigoberto Rios-Esteva, Ph.D.  
Washington State University  
December 2008

Co-Chairs: James M. Lee

Bernd Markus Lange

In this dissertation a combination of mathematical modeling and experimental testing was used to study the regulation of monoterpene metabolism in peppermint. Due to the regulatory complexity and compartmentalization of plant metabolism, it is essential to understand which modeling approach is suitable to tackle a specific scientific problem. Thus, a comprehensive review which summarizes pros and cons of the various modeling approaches is included as the second chapter of this dissertation. For dynamic experimental systems, kinetic modeling is the most reliable strategy, since it accounts for the continuous diurnal, circadian and seasonal changes in the expression and activities of biosynthetic enzymes.

High quality peppermint essential oils are characterized by a complex compositional balance of monoterpenes, with high quantities of (-)-menthol, moderate amounts of (-)-menthone and low levels of (+)-pulegone and (+)-menthofuran. However, under adverse environmental conditions such as low light intensity and drought, the

branch point intermediate (+)-pulegone and the side product (+)-methofuran are accumulated, thus rendering an oil of inferior quality.

Dynamic kinetic modeling of the peppermint monoterpene biosynthetic pathway suggested that competitive inhibition of pulegone reductase, the enzyme catalyzing the conversion (+)-pulegone into (-)-menthone, by (+)-menthofuran, could explain the monoterpene profiles obtained with plants grown under low light conditions. An experimental follow-up study using recombinant pulegone reductase indeed confirmed (+)-menthofuran as a competitive inhibitor of pulegone reductase.

Peppermint plants expressing an antisense version of (+)-menthofuran synthase transcript were shown to contain low levels of (+)-pulegone and (+)-menthofuran. In addition to these desirable effects on essential oil composition, the oil yield in one particular line, designated MFS7, was also enhanced compared to wild-type controls. Based on real-time quantitative PCR assays, the monoterpenoid essential oil composition in MFS7 plants, but not increased yield compared to wild-type controls, could be explained by gene expression patterns. Interestingly, compared to controls, MFS7 plants had a higher density of glandular trichomes, the specialized anatomical structures responsible for the synthesis and storage of essential oils. By integrating several levels of experimental data (glandular trichome distribution, biosynthetic gene expression patterns, and kinetic properties of biosynthetic enzymes) kinetic models accurately simulated experimentally determined monoterpene profiles.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iii
ABSTRACT .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER	
1. General Introduction .....	1
2. Experimental and Mathematical Approaches to Modeling Plant Metabolic Networks .....	7
2.1 Introduction .....	9
2.2 Metabolic Flux Analysis .....	12
2.3 Kinetic Analysis of Metabolic Networks .....	46
2.4 Conclusions .....	61
2.5 References .....	62
3. A Systems Biology Approach Identifies the Biochemical Mechanism Regulating Monoterpene Essential Oil Composition in Peppermint .....	75
3.1 Introduction .....	76
3.2 Results and Discussion .....	80
3.3 Materials and Methods .....	92
3.4 References .....	95
3.5 Supplementary Material .....	99

4. Monoterpenoid Essential Oil Yield in Peppermint is Determined by the Density and Size Distribution of Glandular Trichomes .....	118
4.1 Introduction. ....	119
4.2 Materials y Methods .....	122
4.3 Results and Discussion .....	127
4.4 References .....	142
4.5 Supplementary Material .....	144
5. Conclusion and Future Plans .....	154

## APPENDIX

A. Supplementary Material for the Review Paper “Experimental and Mathematical Approaches to Modeling Plant Metabolic Networks”... ..	158
---	-----

## LIST OF TABLES

Table 2.1. Overview of freely available software packages and online tools for mathematical modeling of plant metabolism. ....	13
Table 2.2 Overview of mathematical modeling papers focusing on plant Metabolism .....	14
Table 4.1 Glandular trichome density and size distribution at 30 d after leaf emergence .....	133



## LIST OF FIGURES

Figure 2.1 Elementary modes analysis of sucrose metabolism in sugar cane culm (based upon data from Rohwer and Botha, 2001) .....	25
Figure 2.2. Metabolic flux analysis of central carbon metabolism in developing soybean embryos (based upon data from Sriram et al., 2004) .....	45
Figure 2.3. Dynamic simulation of changes in metabolite concentrations over time for the metabolic network depicted in Scheme 1 .....	50
Figure 2.4. Determination of flux control coefficients using the direct method of metabolic control analysis. ....	54
Figure 3.1. Outline of <i>p</i> -menthane monoterpene metabolism in peppermint glandular trichomes. ....	78
Figure 3.2. Experimentally determined monoterpene profiles of peppermint plants grown under greenhouse conditions and computer simulation based upon a kinetic mathematical model of mint monoterpene biosynthesis...	84
Figure 3.3. Monoterpene profiles of peppermint plants maintained in growth chambers under low light conditions and computer simulation	

considering a reduction of PR and increase in MFS transcript levels. ....	86
Figure 3.4. Characterization of (+)-menthofuran as a competitive inhibitor of peppermint .....	89
Figure 4.1 Outline of <i>p</i> -menthane monoterpene biosynthesis in peppermint glandular trichomes. ....	121
Figure 4.2. Experimentally determined monoterpene profiles of greenhouse-grown wild-type ( <b>A</b> ), MFS7 transgenic ( <b>B</b> ), and L3H20 transgenic ( <b>C</b> ) peppermint plants. ....	129
Figure 4.3. Experimentally determined monoterpene profiles of wild-type plants grown under various stress conditions: ( <b>A</b> ), low water; ( <b>B</b> ), low light intensity; ( <b>C</b> ), low light intensity and high night temperatures. ....	132
Figure 4.4. Correlation of glandular trichome density and monoterpenoid essential oil yield in peppermint leaves. Greenhouse conditions. ...	138
Figure 4.5. Computer simulation of monoterpene profiles for plants grown at different environmental conditions .....	141

## **Dedication**

*To Dalis, Laura, Camilo and my Family*

## CHAPTER 1

### General Introduction

Essential oils are highly valued as fragrances and flavors. They find commercial application in the food and pharmaceutical industries, with production levels reaching 7000 metric tons per year in the United States (Croteau et al, 2005). One of the most important commercial essential oil plants is peppermint, a perennial herb of the family *Lamiaceae* that produces a complex mixture of p-menthane-type monoterpenes (Lawrence, 1981).

Over the last two decades peppermint has become a model for research on terpenoid biosynthesis and molecular genetics, mainly because of the pioneering molecular biological and biochemical work performed in the laboratory of Dr. Rodney Croteau, an eminent WSU faculty. A wealth of information regarding the genes and enzymes involved in peppermint monoterpene biosynthesis is available (Croteau et al, 2005). However, the regulation of this pathway, particularly under environmental stress conditions, is still poorly understood. Due to its complexity, the monoterpene pathway can not be understood intuitively, thus requiring novel approaches to guide further experimentation aimed at unraveling the regulation of monoterpene biosynthesis.

The integration of mathematical modeling and experimental testing is emerging as a powerful approach for improving our understanding of the regulation of metabolic

pathways. Although being used widely for assessing the control of metabolic flux in microbes, mathematical modeling approaches that require steady-state approximations are of limited utility for understanding complex plant metabolic networks. However, considerable progress has been made when manageable metabolic subsystems were studied (Zimmer et al., 2000; Farquhar et al., 2001; Poolman et al., 2000).

Due to the continuous changes in expression and activity of the enzymes when the plant is exposed to diurnal, circadian and seasonal changes, metabolic modeling of these systems demands dynamic approaches (Morgan and Rhodes, 2002). Yet, a simple dynamic kinetic modeling approach requires mechanistic details and kinetic parameters that in most cases are not available (Giersch, 2000). Most complex mathematical treatments built on biological and biochemical simplifications, parameter estimation, and/or statistical treatments that are computationally demanding. Chapter two of this dissertation entitled “Experimental and mathematical approaches to modeling plant metabolism” contains a comprehensive review of mathematical modeling on plant metabolism. It was intended for explaining, in a language accessible to researchers from different disciplines, diverse and most commonly used mathematical treatments, by means of examples using hypothetical metabolic pathways and examples from the literature. This review chapter summarizes not only different modeling approaches, but also lists and briefly describes the available computational tools.

High quality peppermint oils contain (-)-menthol (Mol) as the principal metabolite, moderate amounts of (-)-menthone, and low levels of (+)-pulegone (PUL)

and (+)-menthofuran. Oil quality and quantity is commonly affected by adverse environmental conditions such as low light intensity or drought, which lead to oil of inferior quality with high amounts of the pathway intermediate PUL, and the side product MF (Burbott and loomis, 1967; Clark and Menary, 1980). It had generally been assumed that peppermint oil yield and composition were primarily controlled at the transcriptional level of the genes involved in monoterpene biosynthetic pathway (McConkey et al., 2000). Metabolic turnover and evaporative losses of oil components were regarded as negligible (Gesherzon et al., 2000).

To study monoterpene metabolism in peppermint (*Mentha piperita*), we used iterative cycles of (dry) mathematical modeling and (wet) laboratory testing. By developing a kinetic mathematical model that accurately describes the behavior of the peppermint monoterpene biosynthetic pathway under various experimental conditions, we were able to generate non-trivial, testable hypotheses regarding poorly understood regulatory mechanisms. Our simulations indicated that the environmentally-regulated changes in monoterpene profiles could only be explained when, in addition to effects on biosynthetic enzyme activities, inhibitory effects of (+)-menthofuran on the branch point enzyme pulegone reductase (PR) were assumed.

Modeling-guided follow-up experiments were used to demonstrate an as yet unidentified role for (+)-menthofuran as a competitive inhibitor of PR. These results indicated that gene expression and posttranslational modulation of enzyme activity were both important factors in regulating peppermint monoterpene biosynthesis. Chapter 3 of

this dissertation entitled “A system biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint” describes the detailed experimental and mathematical treatments used in identifying this inhibitory mechanism.

To further advance our modeling efforts, we identified a need for assessing the effects of additional environmental stresses on the yield and composition of peppermint essential oil. An analogous analysis of transgenic peppermint plants with modulated expression levels of monoterpene biosynthetic genes was also deemed beneficial for increasing the robustness of our modeling and its utility in guiding efforts aimed at modulating essential oil yield and composition. These studies included MFS7, a transgenic line that contains an antisense construct for the suppression of the (+)-menthofuran synthase (MFS) gene, thus leading to significant decreases in the amounts of the undesirable side product (+)-menthofuran (Mahmoud and Croteau, 2001).

During routine analyses of transgenic lines we detected significantly elevated essential oil quantities, when compared to wild-type controls, in the MFS7 antisense line. This finding prompted us to investigate pathway flux-independent mechanisms of oil yield increases. Our data indicate that genotype-dependent and environmental effects on essential oil yield correlate directly with the density and size distribution of glandular trichomes on the leaf surface. Based on these experimental data, the input variables for our existing kinetic mathematical model of peppermint monoterpenoid essential oil biosynthesis were updated. Simulations with this second generation model accurately

predicted the experimentally observed variation of monoterpene yield and composition for antisense and wild-type plants under several environmental conditions. The complete experimental methodology and model simulations can be seen in the Chapter 4 of this dissertation entitled “Monoterpenoid essential oil yield in peppermint is determined by the density and size distribution of glandular trichomes”.

The final Chapter summarizes our results and suggests further improvements to the kinetic model to make it more accurate, reliable and at some point predictive.

## References

Burbott AJ, Loomis WD (1967) Effects of light and temperature on the monoterpenes of peppermint. *Plant Physiol* 42:20-28.

Clark RJ, Menary RC (1980) Environmental effects on peppermint, I. Effect of day length, photon flux density, night temperature and day temperature on the yield and composition of peppermint oil. *Aust J Plant Physiol* 7:685-692.

Farquhar GD, von Caemmerer S and Berry JA, Models of photosynthesis. *Plant Physiol.* 125 (2001), pp. 42–45

Gershenzon J, McConkey ME, Croteau R (2000) Regulation of monoterpene accumulation in leaves of peppermint (*Mentha x piperita* L.). *Plant Physiol* 122:205-214.

Giersch, C., 2000. Mathematical modelling of metabolism. *Current Opinion in Plant Biology* 3, 249-253.

Lawrence BM (1981) *Essential Oils*. Allured Publishing Co., Wheaton IL, pp. 1-81 (McConkey et. al., 2000)

Mahmoud SS, Croteau RB (2003) Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpene reductase. *Proc Natl Acad Sci USA* 100:14481-14486

McConkey ME, Gershenzon J, Croteau R (2000) Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint (*Mentha x piperita* L.). *Plant Physiol* 122:215- 233.



Morgan, J.A., Rhodes, D., 2002. Mathematical modeling of plant metabolic pathways. *Metabolic Engineering* 4, 80-89.

Poolman MG, Fell MA and Thomas S, Modelling photosynthesis and its control. *J. Exp. Bot.* 51 (2000), pp. 319–328

Zimmer W, Brüggemann N, Emeis S, Giersch C, Lehning A, Steinbrecher R and Schnitzler JP, Process-based modelling of isoprene emission by oak leaves. *Plant Cell Environ.* 23 (2000), pp. 585–595

## CHAPTER 2

### **Experimental and mathematical approaches to modeling plant metabolic networks**

Rigoberto Rios-Esteba<sup>1</sup>, Bernd Markus Lange\*

Institute of Biological Chemistry, Center for Integrated Biotechnology, and <sup>1</sup>School of Chemical Engineering and Bioengineering, Washington State University, PO Box 646340, Pullman, WA 99164-6340, USA;

This article is dedicated to the memory of Reinhart Heinrich, one of the founding fathers of metabolic control theory and a pioneer of systems biology.

\* Corresponding author. E-mail address: lange-m@wsu.edu (B.M. Lange).

Reproduced with permission from “Phytochemistry Volume 68, Issues 16-18, August-September 2007, Pages 2351-2374” Dynamic Metabolic Networks  
Copyright © 2007 Elsevier Ltd.

#### **Abstract**

To support their sessile and autotrophic lifestyle higher plants have evolved elaborate networks of metabolic pathways. Dynamic changes in these metabolic networks are among the developmental forces underlying the functional differentiation of organs, tissues and specialized cell types. They are also important in the various interactions of a plant with its environment. Further complexity is added by the extensive compartmentation of the various interconnected metabolic pathways in plants. Thus, although being used widely for assessing the control of metabolic flux in microbes, mathematical modeling approaches that require steady-state approximations are of limited utility for understanding complex plant metabolic networks. However, considerable progress has been made when manageable metabolic subsystems were studied. In this article we will explain in general terms and using simple examples the concepts underlying stoichiometric modeling (metabolic flux analysis and metabolic

pathway analysis) and kinetic approaches to modeling (including metabolic control analysis as a special case). Selected studies demonstrating the prospects of these approaches, or combinations of them, for understanding the control of flux through particular plant pathways are discussed. We argue that iterative cycles of (dry) mathematical modeling and (wet) laboratory testing will become increasingly important for simulating the distribution of flux in plant metabolic networks and deriving rational experimental designs for metabolic engineering efforts.

## **Contents**

1. Introduction
  2. Metabolic flux analysis
    - 2.1. Case study of flux balance analysis
    - 2.2. Metabolic pathway analysis
    - 2.3. Case study of isotope labeling-based metabolic flux analysis
      - 2.3.1. Atom mapping matrices
      - 2.3.2. Isotopomer balancing
      - 2.3.3. Cumomer balancing
      - 2.3.4. Bondomer balancing
    - 2.4. Metabolic flux analysis for modeling plant metabolic networks
  3. Kinetic analysis of metabolic networks
    - 3.1. Case study of kinetic modeling and dynamic simulation
    - 3.2. Kinetic models of plant metabolic networks
    - 3.3. Metabolic control analysis
      - 3.3.1. Case study of metabolic control analysis
      - 3.3.2. Metabolic control analysis for modeling plant metabolic networks
- Acknowledgements
- References

## 2.1 Introduction

Metabolic engineering aims to purposefully alter (using genetic engineering techniques) an organism's metabolic pathways in order to better understand how these pathways work or to redesign them for the production of particular target metabolites. In 2002, the entire January issue of the journal *Metabolic Engineering* (Metabolic Engineering, Volume 4, Issue 1, Pages 1-106, January 2002) was devoted to highlighting the progress, problems and prospects of plant metabolic engineering. Undoubtedly, the plant metabolic engineering community has made considerable progress, which is attested by highly publicized success stories such as the development of provitamin A-producing grains of 'golden rice' (Ye et al., 2000) or the production of thermoplastic polyhydroxyalkanoate polymers in transgenic oilseed rape (Slater et al., 1999). However, further success of plant metabolic engineering at a commercial scale, besides suffering from lack of public acceptance in some parts of the world, has been hampered by an insufficient understanding of the mechanisms controlling flux through plant metabolic pathways, and the outcome of metabolic engineering attempts can not yet be predicted with satisfactory accuracy (DellaPenna, 2001). With the advent of 'omics' technologies (in particular transcriptomics, proteomics, and metabolomics), which allow the global profiling of the abundance of constituents involved in metabolic pathways, our ability to document changes brought about by metabolic engineering is greatly enhanced. However, it has been argued that these approaches have to be augmented by quantitative mathematical modeling, combined in an iterative cycle with experimental testing of

model predictions, to enable a rational design of metabolic engineering strategies (Sweetlove and Fernie, 2005; Fernie et al., 2005; Lange, 2006; Ratcliffe and Shachar-Hill, 2006). Pathway modeling has been an area of strength for chemical and biochemical engineers but most plant biologists are not familiar with the underlying mathematical framework. In this article, we are discussing some the most widely used methods for modeling metabolic networks and illustrate their utility in simplified case studies. We are attempting to provide sufficient information for phytochemists, biochemists, biophysicists and geneticists to understand the basic concepts of mathematical modeling and to allow them to evaluate the available literature. It is our hope that this article will serve as a primer for readers to think about how mathematical modeling can be utilized in providing quantitative descriptions of metabolic network behaviour. Of course, we can not cover the issue in-depth and the reader is referred to the primary literature and several excellent books for further perusal (Heinrich and Schuster, 1996; Fell, 1997; Stephanopoulos et al., 1998).

Various bioinformatic tools are available to compute the topology of genome-scale metabolic networks from experimental data (reviewed in Xia et al., 2004). However, the mathematical modeling methods used today for estimating flux through metabolic pathways (or networks of pathways) require an in-depth knowledge of one or more of the following input data: the (assumed) stoichiometry of all biochemical reactions, reversibility of enzymatic steps, branching patterns of (sub)pathways, uptake rates of a metabolic substrate and conversion rate into end products, subcellular compartmentation, the kinetic constants of the enzymes involved, and possibly profiles of transcript, protein and/or metabolite abundance. Thus, although genome-scale

reconstructions have been attempted for microbes (reviewed in Price et al., 2004), the majority of flux modeling efforts in plants has been restricted to one pathway or a small set of pathways. Mathematical modeling approaches for flux determination can be categorized based on the algorithms, constraints and data types that are utilized, and their selection for a particular project will depend on which prior experimental information is available.

The conceptually simplest method of metabolic flux analysis is the stoichiometric approach, which requires only very limited experimental data; essentially, only the topology of the network, the rate of uptake of a substrate, its conversion rate into products, and the production of biomass are considered. Further developments of this concept, which are based on the use of different algorithms and constraints, are flux balance analysis (FBA), extreme pathway analysis (EPA) and elementary modes analysis (EMA). Because of the limitations of stoichiometric modeling, an experimental approach was developed that is based on labeling experiments with stable isotopes (primarily  $^{13}\text{C}$ ) and subsequent analysis of the distribution of label in different carbon atoms of the target metabolites. Mechanistic (kinetic) models incorporate knowledge regarding the kinetic properties of enzymes involved in a metabolic network and can be used to simulate how intracellular fluxes will change when the experimental system is perturbed. Metabolic control analysis (MCA) has become the most widely used kinetics-based tool to gain a quantitative understanding of metabolic networks. In the following paragraphs we will discuss the different mathematical modeling approaches and will highlight progress made in applying these different approaches to understanding the control of plant metabolism. In Table 2.1 an overview of freely available software packages and online tools relevant

to the mathematical modeling of plant pathways is presented. A critical assessment of these tools is beyond the scope of this article; for an excellent article evaluating tools for kinetic modeling the reader is referred to Alves et al. (2006). Since we will not be able to discuss every paper that has appeared in the field of mathematical modeling of plant pathways in this review article, we are providing a comprehensive listing (sorted by modeling approach and scientific area) in Table 2.2 and focus on the discussion of selected articles in the text.

## **2.2. Metabolic flux analysis**

In Metabolic Flux Analysis (MFA) intracellular fluxes are calculated by using a stoichiometric model for the major intracellular reactions and applying mass balances around intracellular metabolites. This concept is based upon the fundamental law of mass conservation (all mass inputs into a metabolic network must be recovered, transformed or remain unchanged; Heinrich and Schuster, 1996). We distinguish between purely stoichiometric approaches (only a limited number of metabolite measurements are made; discussed in 2.1. and 2.2.) from those that utilize isotope labeling experiments (discussed in 2.3.).

Ordinarily, the number of reactions (fluxes) is greater than the number of intracellular metabolites and the solution of metabolite balancing equations (a system of ordinary differential equations (ODEs) used as a mathematical framework to determine flux) will allow an infinite number of possible solutions (combination of fluxes).

**Table 2.1.** Overview of freely available software packages and online tools for mathematical modeling of plant metabolism.

Modeling method	Software tool	References
<b>Isotope labeling-based metabolic flux</b>	13C-Flux	Wiechert et al. (2001)
	FiatFlux	Zamboni et al. (2005)
	Mathematica-based program (no name given)	Selivanov et al. (2006)
	NMR2Flux	Sriram et al. (2004)
<b>Elementary modes analysis</b>	FluxAnalyzer	Klamt et al. (2003)
	SNA	Urbanczik (2006)
	YANA	Schwarz et al. (2005)
<b>Extreme pathway analysis</b>	FluxAnalyzer	Klamt et al. (2003)
	Expa	Bell and Palsson (2005)
<b>Kinetic modeling</b>	CellDesigner (latest version is 3.5.1)	Funahashi, et al. (2003)
	Cellware (latest version is 3.0.2)	Dhar et al. (2004)
	COPASI (latest version is 4.0)	Hoops et al. (2006)
	Dizzy (latest version is 1.11.4)	Ramsey et al. (2005)
	Dynetica (latest version is 2.0beta)	You et al. (2003)
	E-CELL 2	Takahashi et al. (2003)
	GEPASI (latest version 3.30)	Mendes (1993)
	JDesigner/Jamac (latest version is 2.0.35)	<a href="http://sbw.kgi.edu/">http://sbw.kgi.edu/</a>
	JSim	<a href="http://nsr.bioeng.washington.edu/PLN/Software">http://nsr.bioeng.washington.edu/PLN/Software</a>
	JWS Online	Olivier and Snoep (2004)
	METATOOL 5.0	von Kamp and Schuster (2006)
	PLAS (Power Law Analysis and Simulation)	<a href="http://www.dqb.fc.ul.pt/docentes/aferreira/plas.html">http://www.dqb.fc.ul.pt/docentes/aferreira/plas.html</a>
	PySCeS	Olivier et al. (2005)
	Systems Biology Toolbox for MATLAB	Schmidt and Jirstrand (2005)
	Vcell (latest version is 4.2)	<a href="http://www.nrcam.uchc.edu/">http://www.nrcam.uchc.edu/</a>
WebCell (latest version is 3.0)	Lee et al. (2006)	
<b>Reporting standards and repositories</b>	BioModels	Le Novere et al. (2006)
	MFAML	Yun et al. (2005)
	MIRIAM	Le Novere et al. (2005)
	JWS Online	Olivier and Snoep (2004)
	SBML	SBML Forum (2003)
	WebCell (latest version is 3.0)	Lee et al. (2006)



**Table 2.2** Overview of mathematical modeling papers focusing on plant metabolism.

Modeling method	Pathway	Comment	References
<b>Isotope labeling-based metabolic flux analysis</b>	Central carbon metabolism	Quantification of compartmented metabolic fluxes in maize root tips.	Dieuaide-Noubhani et al. (1995)
		In vivo pyruvate synthesis in maize roots was evaluated using a precursor-product.	Edwards et al. (1998)
		Heterotrophic tobacco callus lines expressing a rat liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase were used to calculate rates of glucose metabolism.	Fernie et al. (2001)
		Calculation of 18 fluxes of central metabolism based on (13)C enrichments in tomato suspension cultures.	Rontein et al. (2002)
		Analysis of central metabolism in Brassica napus embryos based on (13C) enrichments in storage lipids and proteins.	Schwender and Ohlrogge (2002)
		Quantitative metabolic flux model for the reaction network of glycolysis and the oxidative pentose phosphate pathway in Brassica napus embryos.	Schwender et al. (2003)
		Based on isotopomer abundances in soybean seed storage protein and starch hydrolysates, fluxes through pathways of central carbon metabolism were calculated.	Sriram et al. (2004)
		Flux contributions via several pathways of central carbon metabolism were estimated based on feeding labeled glucose to tobacco.	Ettenhuber et al. (2005a)
		Relative flux contributions by different pathways of primary metabolism in maize kernels were determined by simulation of the isotopolog space of glucose.	Ettenhuber et al. (2005b)
		A new futile cycle (glucose phosphate to glucose) was discovered that consumes about 40 % of all ATP generated in maize root tips.	Alonso et al. (2005)
		Metabolism in tuber discs (response to low temperature) was monitored by determining the redistribution of radiolabel following incubation in [U-(14)C]glucose.	Malone et al. (2006)
		(13)C-labeling experiments using glucose were carried out with kernels of maize inbred lines, heterotic hybrids, and starch-deficient mutants.	Spielbauer et al. (2006)
		Choline metabolism	Conversion of [(33)P]phospho-EA, [(33)P]phospho-monomethylethanolamine, or [(14)C]formate into choline in tobacco leaf disks.
Definition of constraints on glycine betaine syntheses based on [(14)C]choline labeling experiments and in vivo (31)P NMR analyses of tobacco leaf disks.	McNeil et al. (2000b)		
Mitochondrial metabolism	Mitochondrial fluxes were determined after labeling Brassica napus embryos with (13C)glucoses, (13C)alanine, (13C)glutamine,	Schwender et al. (2006)	
Phenylpropanoid metabolism	Modulation of phenylpropanoid metabolism in wound-healing potato tuber tissue.	Matsuda et al. (2003)	
	Investigation of the metabolic pathways in Petunia hybrida petals leading from phenylalanine to benzenoid compounds.	Boatright et al. (2004)	
	Analysis of the effects of beta-1,3-oligosaccharide elicitor on the metabolism of phenylpropanoids in potato tuber.	Matsuda et al. (2005)	
<b>Elementary modes analysis</b>	C3 photosynthesis	Analysis of photosynthate metabolism in the chloroplast.	Poolman et al. (2003)
	Central carbon metabolism	All possible routes of futile cycling of sucrose in sugar cane culm were enumerated the available kinetic data for the pathway enzymes were then used a kinetic model.	Rohwer and Botha (2001)
		Measurements of mass balance, Rubisco enzyme activity, stable isotope labelling and analysis of elementary flux modes (Brassica napus embryos).	Schwender et al. (2004b)

Table 2.2 (continued)

Modeling method	Pathway	Comment	References
Kinetic modeling	C3 photosynthesis	Biochemical model of photosynthetic CO <sub>2</sub> assimilation in leaves of C3 species.	Farquhar et al. (1980)
		Dynamic model of photosynthesis.	Gross et al. (1991)
		Sensitivity of C3 photosynthesis to increasing CO <sub>2</sub> concentration.	Kirschbaum (1994)
		Modeling of photosynthesis in fluctuating light (incl. stomatal conductance, biochemical activation and pools of key photosynthetic intermediates).	Kirschbaum et al. (1997)
		Estimation of carbon gain in sunfleck light regimes.	Pearcy et al. (1997)
		Improvement of a previous kinetic model for the Calvin cycle and starch production.	Pettersson (1997)
		Photosynthetic carbohydrate formation in C3 plants under conditions of light and carbon dioxide saturation.	Pettersson and Ryde-Pettersson (1988)
		Flux control of the malate valve in leaf cells.	Fridlyand et al. (1998)
		Oscillations in photosynthesis caused by changes in the ATP/ADP ratio or modulation of the pH gradient across the plastidial envelope membrane.	Fridlyand (1998)
	Regulation of the Calvin cycle.	Fridlyand and Scheibe (1999)	
	C4 photosynthesis	Prediction of net photosynthesis and stomatal conductance from leaves of C4 plants.	Collatz et al. (1992)
		Mechanistic leaf photosynthesis model for C4 grasses.	Chen et al. (1994)
		Experimental data using <i>Zea mays</i> at different stages of development were analyzed using mathematical models of C4 photosynthesis.	He and Edwards (1996)
	CAM photosynthesis	Computer model comprising light reactions in PS II and PS I, electron-proton transport reactions in mesophyll and bundle sheath chloroplasts.	Laisk and Edwards (2000)
		Simulation of Crassulacean acid metabolism (CAM). Model of crassulacean acid metabolism (CAM) describing the varying concentrations of pools of major metabolites by a system of coupled nonlinear differential equations is proposed.	Nungesser et al. (1984) Blasius et al. (1997)
Hysteresis switch in an oscillatory model of Crassulacean acid metabolism.		Neff et al. (1998)	
Central carbon metabolism	Oscillatory model of crassulacean acid metabolism (CAM) describing the CO <sub>2</sub> uptake and nocturnal acidification of CAM plants.	Blasiuset al. (1998)	
	Non-linear theoretical model of mechanism of endogenous circadian photosynthesis oscillations of plants performing crassulacean acid metabolism (CAM).	Blasius et al. (1999)	
	Regulation of several enzymes in central carbon metabolism as a response to hypoxia in maize root tips.	Roscher et al. (1998)	
Carotenoid metabolism	Kinetic model of the branchpoint between methionine and threonine biosynthesis	Curien et al. (2003)	
	Analysis of central carbon metabolism in <i>Catharanthus roseus</i> hairy root cultures.	Leduc et al. (2006)	
	Kinetic model of the xanthophyll cycle.	Sielewiesiuk and Gruszecki (1991)	
		Kinetics of the two de-epoxidation steps occurring in the xanthophyll cycle.	Latowski et al. (2000)

Table 2.2 (continued)

Modeling method	Pathway	Comment	References
<b>Kinetic modeling (continued)</b>	Choline metabolism	Modeling of choline metabolism in transgenic plants.	McNeil et al. (2000b)
		Modeling of choline metabolism in transgenic plants.	Nuccio et al. (2000)
	Isoprene emission	Model describing the formation of isoprene in oak under varying environmental conditions.	Zimmer et al. (2000)
	Mitochondrial respiration	Modeling of the respiratory network in plant mitochondria.	Krab K (1995)
	Redox regulation	Dissection of the superoxide dismutase-ascorbate-glutathione-pathway in chloroplasts.	Polle (2001)
<b>Metabolic control analysis</b>	C3 photosynthesis	Simulation of dynamic and steady-state behaviour of the Calvin cycle reactions of the chloroplast, including starch synthesis and degradation, and triose phosphate export.	Poolman et al. (2000)
		Regulation of fluxes in transgenic plants with reduced activities of Calvin-cycle enzymes.	Fridlyand and Scheibe (2000)
	Central carbon metabolism	Analysis of mutants of phosphoglucose isomerase in the cytosol and chloroplast of <i>Clarkia xantiana</i> .	Kruckeberg et al. (1989)
		Analysis of glycolysis in aged disks of tuber tissue from transgenic potatoes expressing different amounts of phosphofructokinase.	Thomas et al. (1997a)
		Distribution of glycolytic flux control between the steps of glycolysis in aged disks of potato tuber under aerobic conditions.	Thomas et al. (1997b)
		Control of gluconeogenesis in endosperm from 4-day-old castor bean seedlings.	Runquist and Kruger (1999)
		Kinetic model of the branchpoint between methionine and threonine biosynthesis	Curien et al. (2003)
	Biopolymer production	Prospects of producing the copolymer poly-(3-hydroxybutyrate-co-3-hydroxyvalerate) in plant plastids.	Daee et al. (1999)
	Carotenoid metabolism	Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner.	Fraser et al. (2002)
	Glutamine synthetase/ glutamate synthase cycle	Metabolic control analysis of the glutamine synthetase/glutamate synthase cycle of barley chloroplasts.	Baron et al. (1994)
	Lipid metabolism	The importance of acetyl-CoA carboxylase in regulation of lipid synthesis for barley and maize leaves was quantitatively assessed using inhibitor titration studies.	Page et al. (1994)
		Analysis of de novo fatty acid biosynthesis in developing sunflower seeds.	Martinez-Force et al. (2000)
		Study of consequences of omega -6-oleate desaturase deficiency on mitochondrial membrane function in <i>Arabidopsis thaliana</i> .	Caiveau et al. (2001)
		Analysis of lipid biosynthesis in tissue cultures from oil crops (olive, oil palm).	Ramli et al. (2002)
	Analysis of lipid biosynthesis in tissue cultures from oil crops (olive, oil palm).	Ramli et al. (2005)	
Redox regulation	Analysis of glutathione (GSH) and phytochelatin synthesis.	Mendoza-Cozatl and Moreno-Sanchez (2006)	
Mitochondrial respiration	Analysis of the control of phosphorylation-coupled respiration in isolated plant mitochondria.	Padovan et al. (1989)	

This is referred to as an under-determined system. Additional experimental measurements and derived constraints can be used to reduce the number of unknown fluxes. As a final step it is possible to introduce optimization criteria (e.g., the network strives to produce a certain essential metabolite at highest possible concentrations), which allows linear programming (a well-developed mathematical approach for solving diverse optimization problems) to be applied for solving the system of ODEs.

### 2.2.1. Case study of flux balance analysis

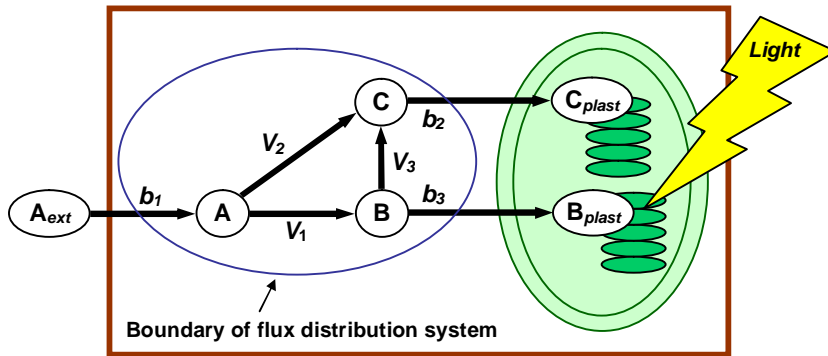
Stoichiometric analyses require knowledge about the (assumed) topological structure of the metabolic network under consideration and assume that the network operates at a pseudo-steady state (the macroscopic variables – flux and metabolite concentrations – change only to a tolerable extent over a specific time span) (Heinrich and Schuster, 1996; Schilling et al., 1999; Klamt and Stelling, 2003). For the simultaneous solution of mass balance equations, a homogeneous system of linear algebraic equations is set up to define a vector of metabolite concentrations  $\mathbf{M}$ , a stoichiometric matrix  $\mathbf{S}$ , of order  $m \cdot n$  ( $\mathbf{m}$ , number of pathway metabolites;  $\mathbf{n}$ , number of metabolic fluxes), and a vector  $\mathbf{v}$  of net reaction rates. The variation of a metabolite concentration over time ( $dM/dt$ ), is proportional to the rate of reaction at which it is synthesized minus the rate of reaction at which it is consumed. At steady state conditions the following expression is obtained:

$$\frac{dM}{dt} = \mathbf{S} \cdot \mathbf{v} = 0$$

We will now use a simple example to illustrate the use of flux balance analysis (FBA).

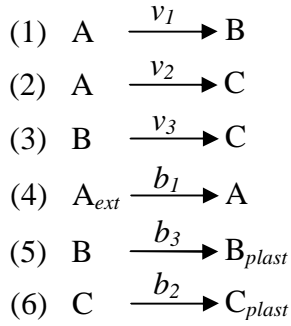
Let us consider the metabolic network defined in Scheme 1.

**Scheme 1.** Example network to illustrate flux balance analysis.



A plant cell imports the metabolite  $A_{ext}$  from the apoplastic space (the internal metabolite is referred to as  $A$ ). The metabolic network consists of two reactions that result in the formation of metabolites  $B$  and  $C$ . These cytosolically synthesized metabolites are then transported to plastids (they are now referred to as  $B_{plast}$  and  $C_{plast}$ ) and associate with thylakoids. Our purpose is to determine the internal flux distribution, using FBA, in the subnetwork circumscribed by the blue oval of Scheme 1.

The metabolic reactions and transport processes can be summarized as follows ( $v_i$  are reaction rates,  $b_i$  are transport rates):



The mass balance of a metabolite (variation over time  $dM/dt$ ) is defined as the difference between the rate(s) of production and the rate(s) of consumption. Applied to all metabolites in our network, the following set of coupled ODEs can be formulated:

**Scheme 2.** Dynamic mass balance equations.

$$\frac{dA}{dt} = b_1 - v_1 - v_2$$

$$\frac{dB}{dt} = v_1 - b_3 - v_3$$

$$\frac{dC}{dt} = v_2 + v_3 - b_2$$

We have now defined the metabolic network and have derived the corresponding dynamic mass balance equations, which can also be rewritten in matrix notation (Scheme 3). To calculate the flux distribution in the network under consideration, we need to identify imposed restrictions (constraints) to the system. If we consider that the macroscopic variables (metabolite concentrations and fluxes) do not change considerably over a certain time span (steady-state assumption), the dynamic mass balance equations can be represented as  $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$ , where  $\mathbf{S}$  is the stoichiometric matrix and  $\mathbf{v}$  the flux vector:

**Scheme 3.** Mass balance equations in matrix form.

$$\frac{d}{dt} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The arrows in Scheme 3 indicate which flux in the flux vector corresponds to which stoichiometric coefficient in the first line of the stoichiometric matrix based on the first equation (specifying  $dA/dt$ ) in Scheme 2. Lines 2 and 3 of the stoichiometric matrix are obtained in the same way for  $dB/dt$  and  $dC/dt$ , respectively. For our example the number of unknown fluxes ( $v_1$ ,  $v_2$ ,  $v_3$ ,  $b_1$ ,  $b_2$ , and  $b_3$ ) is greater than the number of metabolites ( $A$ ,  $B$ , and  $C$ ), which means that the steady-state solution of fluxes is under-determined. Thus, additional constraints such as measurable metabolite concentrations and measurable fluxes can be used to uniquely determine the flux distribution. Let us assume that we have measured the concentrations of metabolites  $A_{ext}$  and  $A$  over a certain period of time; we then calculate the uptake rate  $b_1$  (for this example we assume this rate is in the range of up to  $10 \text{ pmol (g fresh weight} \cdot \text{s)}^{-1}$ ). Let us also assume that we have measured the concentrations of metabolites  $B_{plast}$  and  $C_{plast}$  and we observed that their ratio under various conditions is always greater than or equal to 3 : 1. Further assumptions are that the cell under consideration is experiencing high light exposure and that compound  $B_{plast}$  is an essential component of the high light response of plants, which should be synthesized at high levels under these conditions. In order to determine the internal flux distribution, we can thus formulate an objective function  $Z$  (the cell

maximizes the production of  $B_{plast}$  and, because  $B_{plast}$  and  $C_{plast}$  are linked as defined above, also  $C_{plast}$ ), which can be treated as a linear optimization problem:

$$\text{Maximize } Z = B_{plast} + C_{plast} = b_2 + b_3 \quad \text{Objective function}$$

Summary of constraints:

- (1)  $S \cdot v = 0$       *Mass balance constraint assuming steady-state*
- (2)  $b_1 \leq 10$       *Inferred by measurement of concentrations of  $A_{ext}$  and  $A$*
- (3)  $b_3 \geq 3 \cdot b_2$       *Inferred by measurement of concentrations of  $B_{plast}$  and  $C_{plast}$*
- (4)  $v_1, v_2, v_3, b_1, b_2, b_3 \geq 0$       *Assumption that all reactions and the direction of transport are not reversible*

Optimization problems with an objective function are oftentimes solved using an approach called linear programming. This method identifies one solution (in this case a maximum) that satisfies all constraints. One popular technique for the numerical solution of a linear programming problem is the simplex algorithm. Because of space limitations we can not discuss all iterations of this process in this review article, but for those interested, the entire solution for our example is shown in Supplementary Material 1. The following optimal solution for the problem under consideration (fluxes are expressed in  $\text{pmol (g fresh weight} \cdot \text{s)}^{-1}$ ) was obtained:

$$b_1 = 10; b_2 = 2.5; b_3 = 7.5; v_1 = 7.5; v_2 = 2.5; v_3 = 0$$



Optimizations for determining fluxes through metabolic pathways even of limited complexity (as in our example) are quite tedious. If a larger number of reactions is considered (or a network of pathways), the solution has to be determined computationally.

### ***Limitations of stoichiometric analyses***

The key to a successful stoichiometric analysis is the judicious selection of constraints but its solution demands the experimental determination of unknown intracellular fluxes which might be difficult to quantify without further simplifications, restricting the reliability of the obtained fluxes. However, even if for a particular metabolic pathway the influx of a substrate and the efflux of an end product are known, there are several cases when stoichiometric MFA can not predict flux distribution accurately: if (1) reversible reactions have to be considered, (2) parallel pathways exist or pathways occur in more than one compartment, (3) metabolic cycles are part of the metabolic network, (4) enzyme cofactors are not balanced, (5) pathways are compartmentalized, and/or (6) the steady-state assumption does not hold because of diurnal, seasonal or environmental fluctuations (Stephanopoulos et al., 1998; Varner and Ramkrishna, 1999; Wiechert, 2001). It is not unusual that one or more of these conditions apply in plants and that results with stoichiometric analyses are at best ambiguous (Roscher et al., 2000; Morgan and Rhodes, 2002). So why would we discuss FBA in such detail when its utility for modeling plant pathways (and networks of pathways) is very limited? In the upcoming paragraphs we will introduce other mathematical modeling approaches that use the same mathematical framework

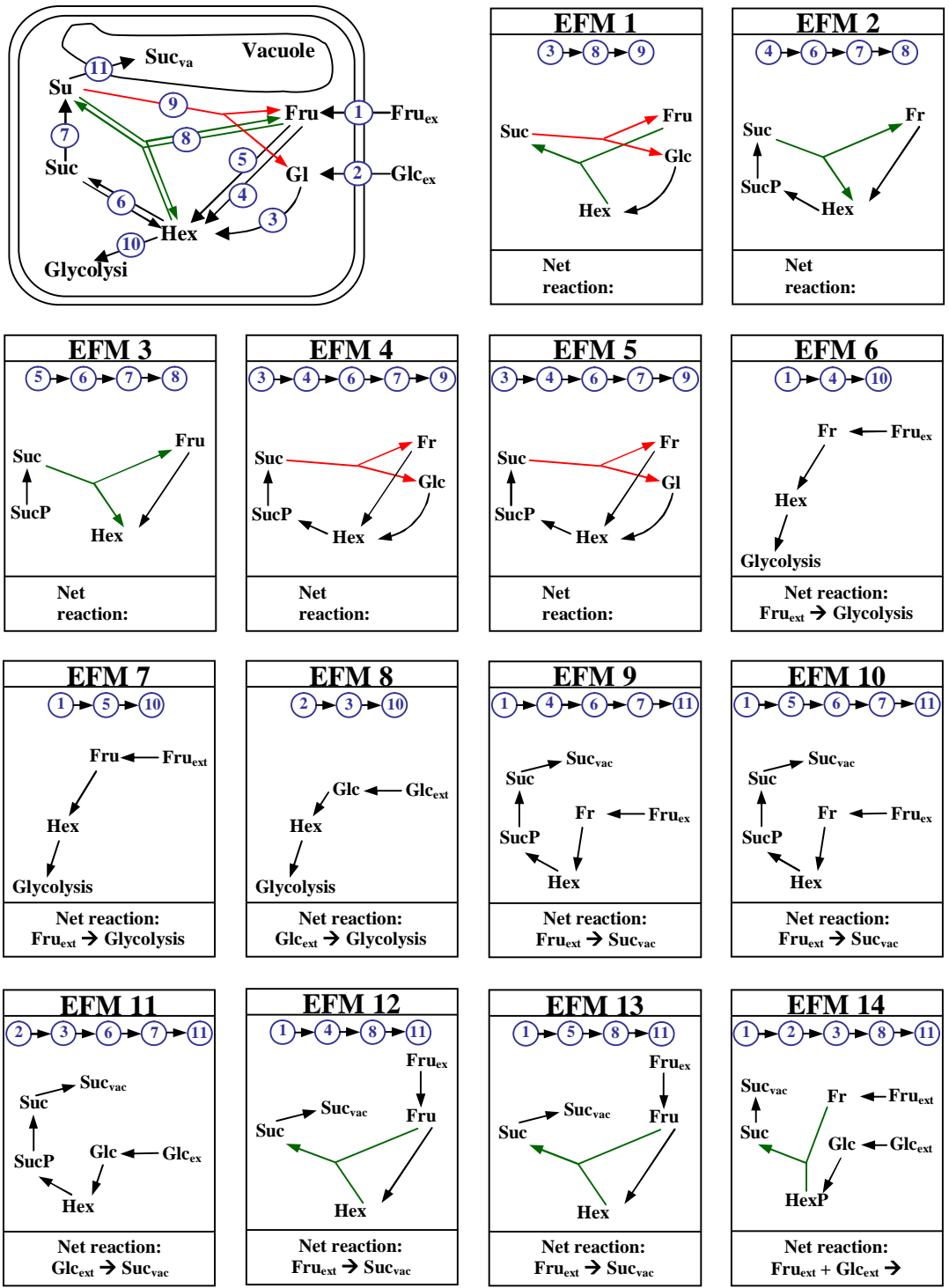
(constraint-based models using reaction stoichiometry) and in this article FBA serves as an introductory example of a balancing problem.

### **2.2.2. Metabolic pathway analysis**

#### **Elementary Modes Analysis**

FBA assumes a certain topology for the metabolic network under consideration. In cases where numerous possibilities for flux distribution exist (e.g., branched pathways), FBA does not lead to unambiguous solutions. Several approaches for metabolic pathway analysis (MPA) have been developed that also build on constructing models based upon reaction stoichiometries but help to find unique solutions to determine flux distribution in metabolic networks. One of these methods is elementary modes analysis (EMA), in which elementary flux modes define the minimum set of reactions that can operate at steady-state under certain constraints and that can not be further decomposed conceptually (Schuster and Hilgetag, 1994; Schuster et al., 1999; Stelling et al., 2002; Gagneur and Klamt, 2004). A complete and thus unique set of pathways, which represents all possible solutions that meet the constraints imposed on the metabolic network under consideration, is generated. The computational challenge that arises from such network decompositions is that, even for relatively simple pathways, thousands of elementary flux modes may have to be considered, particularly when nodes (metabolites) with high connectivities (involvement in numerous reactions) need to be considered. Stelling et al. (2002) introduced an approach to calculate ‘control-effective fluxes’, which assigns an ‘efficiency’ to each elementary mode to relate its output (enhanced growth,

production of energy equivalents, increased photosynthetic rate) to the bioenergetic investment required (synthesis of enzymes). Then, the determined fluxes are weighted by each mode's efficiency, which leads to the generation of control-effective fluxes. Rohwer and Botha (2001) used EMA to outline viable pathways for the futile cycling of sucrose in sugar cane culm (Figure 2.1). Earlier feeding experiments with labeled isotopes had shown that significant futile cycling occurred in sucrose-accumulating sugar cane, which was hypothesized to be a limiting factor in sucrose accumulation (Whittaker and Botha, 1997). Interestingly, EMA revealed that the first 5 of the 14 modes shown in Figure 2.1 represented futile cycles, which do not result in the accumulation of an end product (Rohwer and Botha, 2001). Modes 6 – 8 exemplify the utilization of hexoses to provide precursors for glycolysis. Modes 9, 10, 12 and 13 lead to the accumulation of sucrose in vacuoles using fructose as a precursor, whereas mode 11 is based upon the metabolization of glucose. Mode 14 takes into account that both fructose and glucose can be used as precursors for sucrose biosynthesis. The information regarding possible pathways was then integrated with data obtained using a modeling approach that accounts for the kinetic and thermodynamic properties of the enzymes involved in the sucrose accumulation network (for details regarding kinetic modeling see section 2.3 of this review).



**Figure 2.1.** Elementary modes analysis of sucrose metabolism in sugar cane culm (based upon data from Rohwer and Botha, 2001). Abbreviations: ext, external; Fru, fructose; Glc, glucose; HexP, hexose phosphates; Suc, sucrose, SucP, sucrose phosphate; vac, vacuolar. For details see text.

Based on this integrative analysis the authors suggested a metabolic engineering strategy for enhancing sucrose accumulation by overexpression of either (1) the fructose importer, (2) the glucose importer, or (3) the vacuolar sucrose importer (in Figure 2.1, these correspond to enzymes 1, 2 and 11, respectively). In addition to modulating enzymes involved in transport, the kinetic modeling also suggested that a reduction in the activity of invertase (enzyme 9 in Figure 2.1) should be a viable strategy for increasing sucrose levels (Rohwer and Botha, 2001). The combination of EMA and kinetic modeling has allowed the authors to formulate hypotheses that can now be tested experimentally. Poolman and colleagues (2003) applied EMA to investigate feasible pathways of central carbon metabolism in the chloroplast stroma (in particular the Calvin cycle, the oxidative pentose phosphate pathway and triose phosphate transport) under different light conditions. Based upon these analyses, Calvin cycle reactions, the oxidative pentose phosphate pathway and the thioredoxin system are combined to break down transitory starch for the synthesis of triose phosphates. The authors also propose that in the dark the oxidative pentose phosphate pathway operates cyclically and leads to the formation of C3, C4 and C5 but, unexpectedly, not C6 sugar phosphates. Schwender and colleagues (2004b) performed an EMA analysis of oil biosynthesis (from glucose to fatty acids) in developing oilseed rape embryos to evaluate the contribution of various alternative pathways. In combination with isotope labeling experiments and enzyme activity measurements, these studies demonstrated that carbon fixation in these seeds operates in the absence of a Calvin cycle, which was shown to increase the carbon use efficiency by combining the activity of Rubisco with the non-oxidative reactions of the pentose phosphate pathway. Compared to a conversion of hexose phosphates through

glycolysis this newly discovered route increases the acetyl-CoA production from hexose phosphates by 20 % and reduces the loss of CO<sub>2</sub> (produced by pyruvate dehydrogenase) by 40 %. This might explain why developing seeds of numerous plants have high Rubisco but negligible Calvin cycle activities.

### **Extreme Pathway Analysis**

Another approach for MPA utilizes the concept of convex analysis, which identifies a unique set of ‘extreme pathways’ that define the topology of the metabolic network under consideration (Schilling et al., 2000; Schilling et al., 2001; Papin et al., 2002; Price et al., 2003). Every steady-state flux can be expressed as a non-negative linear combination of these extreme pathways. A limitation of extreme pathway analysis (EPA) relates to the fact that the number of extreme pathways increases exponentially with the size of the network under consideration. Thus, this approach has been applied to networks of reduced size (division into subsystems) or complexity (focus on a particular process altered by an experimental treatment) (reviewed in Price et al., 2003). Network analyses to evaluate the topological properties of plant metabolic networks have been performed for metabolic pathways in plastids (Wang et al., 2006) and it is conceivable that EPA could be employed successfully for specific pathways (e.g., the Calvin cycle and its branchpoints leading into central carbon metabolism) within such a network. In fact it would be quite interesting to assess the outcome of such an analysis in light of the published studies using EMA.

### **2.2.3. Case study of isotope labeling-based metabolic flux analysis**

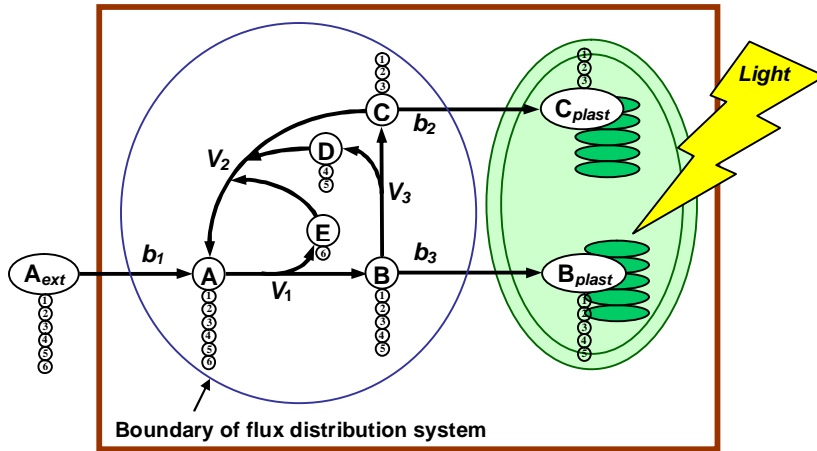
Steady-state isotope labeling-based MFA has been employed successfully when reversible enzymatic steps, metabolic cycles and/or subcellular compartmentation have to be considered for a particular metabolic network of interest (for recent reviews see Roscher et al., 2000; Giersch, 2000; Wiechert, 2001; Morgan and Rhodes, 2002; Schwender et al., 2004; Wiechert and Nöh, 2005; Ratcliffe and Shachar-Hill, 2006). The choice of the labeled precursor is of paramount importance as feeding with different isotopomers (metabolites containing an isotopic atom in different positions) may lead to differences in the labeling patterns of the constituents of a metabolic network. The isotopically labeled precursor is supplied continuously until the pool sizes and labeling patterns of all metabolites remain essentially constant (metabolic and isotopic steady-state). The proportion of isotope-labeled atoms in each metabolite (fractional enrichment) can then be quantified using radiotracer analysis (if radioisotopes are used), nuclear magnetic resonance (NMR; if isotopes such as  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  or  $^{31}\text{P}$  are used) or analytical techniques based on mass spectrometry (MS; applicable for all commonly used isotopes) (for reviews see Szyperski, 1998; Ratcliffe and Shachar-Hill, 2001). The advantages and disadvantages of using a particular technology for assessing fractional isotope enrichments are discussed in-depth in Szyperski (1998). Experimentally, fractional enrichments can be achieved using various different methods: (1) the substrate is isotope-labeled in one (or more) specific position(s) and the isotopic enrichment in cellular metabolite pools is monitored (positional labeling); (2) two different substrates,

of which only one is isotope-labeled, are proffered and isotope enrichments are quantified (dilution of isotope enrichment); (3) a mixture of unlabeled and uniformly isotope-labeled substrate are supplied and the metabolic flux information is contained in the integrity of bonds formed (bond labeling); and (4) a combination of the above techniques (for more details see Roscher et al., 2000).

To illustrate the analysis of fractional enrichments and the estimation of metabolic fluxes, we will use a metabolic network very similar to that depicted in Scheme 1, with the difference that the reaction of  $v_2$  occurs in the opposite direction, thus resulting in a futile cycle with  $v_1$ ,  $v_2$  and  $v_3$  (Scheme 4).  $A_{ext}$  represents a six-carbon metabolite that is proffered as a mixture in unlabeled and U- $^{13}\text{C}$ -labeled form. After uptake of  $A_{ext}$  into the cell,  $A$  is converted to the five-carbon metabolite  $B$  and the one-carbon metabolite  $E$ .  $B$  is then cleaved to form the three-carbon metabolite  $C$  and the two-carbon metabolite  $D$ . Both  $B$  and  $C$  can be transported to chloroplasts. Alternatively, the three-carbon metabolite  $C$  can react with  $D$  and  $E$  to complete a metabolic cycle and regenerate  $A$ .

**Scheme 4.** Example network to illustrate the isotopomer and cumomer approaches to metabolic flux analysis.





### 2.2.3.1. Atom mapping matrices

First, we are going to discuss the steps to derive the principal equations for modeling isotope distributions which can then be used to estimate metabolic fluxes. An important advance was the introduction of the concept of atom mapping matrices (AMMs), which describe the transfer of atoms from reactants to products, thus resulting in a mapping matrix for each reactant-product pair (Zupke and Stephanopoulos, 1994). This method starts with representing the degree of isotope enrichment of each metabolite in vector form (label distribution vector) (note that in the literature the terms ‘metabolite vector’ and ‘metabolite specific activity vector’ are also used):

$$A = \begin{bmatrix} A(1) \\ A(2) \\ A(3) \\ A(4) \\ A(5) \\ A(6) \end{bmatrix} \quad B = \begin{bmatrix} B(1) \\ B(2) \\ B(3) \\ B(4) \\ B(5) \end{bmatrix} \quad C = \begin{bmatrix} C(1) \\ C(2) \\ C(3) \end{bmatrix} \quad \text{Label distribution vector}$$

Because we are proposing the use of a mixture of unlabeled and  $^{13}\text{C}$ -labeled substrate in our hypothetical experiment, the three-carbon metabolite  $C$  can have 8 different isotope labeling states ( $2^n$  isotopomers, where  $n$  is the number of carbons in the metabolite):

**Scheme 5.** Example showing all possible isotopomers of a three-carbon metabolite. Unlabeled atoms are depicted as hollow circles, whereas isotope-labeled atoms are shown as gray circles. The lower panel depicts labeling pattern formulated in vector notation.

								<b>Isotopomer</b>							
								<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
①	①	①	①	①	①	①	①	①	①	①	①	①	①	①	
②	②	②	②	②	②	②	②	②	②	②	②	②	②	②	
③	③	③	③	③	③	③	③	③	③	③	③	③	③	③	
$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	

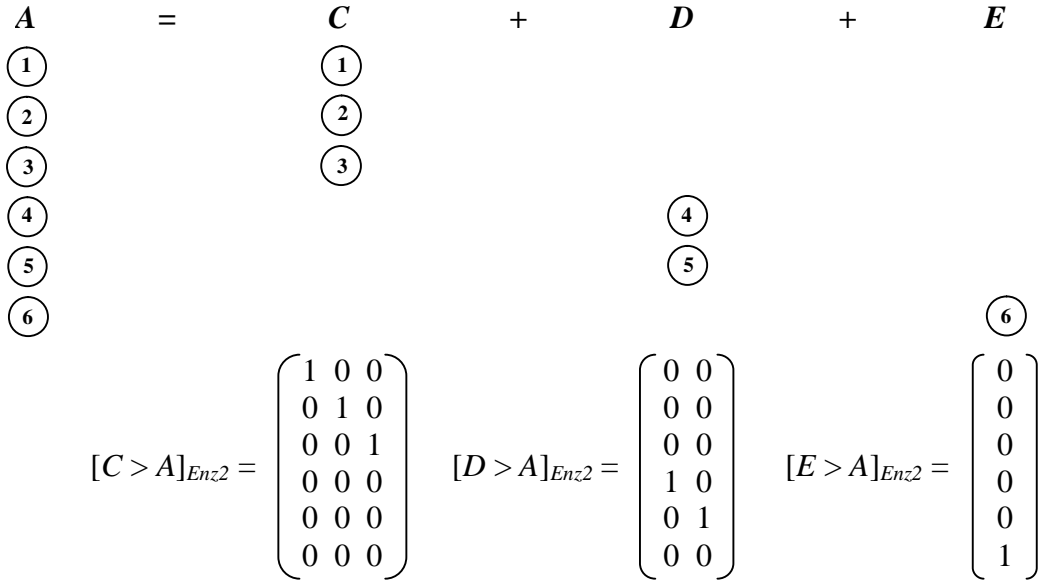
The second step is then to derive the AMMs. For the metabolic network shown in Scheme 4, the following processes are considered when constructing AMMs:

**Scheme 6.** Construction of atom mapping matrices.

<b>Reaction</b>	<b>Function / Notation</b>
(1) $\longrightarrow A$	<i>Transport</i>
(2) $\xrightarrow{Enz1} B + E$	$[A > B]_{Enz1}; [A > E]_{Enz1}$
(3) $\longrightarrow B_{plast}$	<i>Transport</i>
(4) $B \xrightarrow{Enz3} C + D$	$[B > C]_{Enz3}; [B > D]_{Enz3}$
(5) $C \longrightarrow C_{plast}$	<i>Transport</i>
(6) $C + D + E \xrightarrow{Enz2} A$	$[C > A]_{Enz2}; [D > A]_{Enz2}; [E > A]_{Enz2}$

We will now use reaction (6) from Scheme 6 as an example for constructing AMMs. In this notation  $[C > A]_{Enz2}$  represents an atom mapping matrix that describes the transfer of carbon from metabolite *C* to metabolite *A*, catalyzed by enzyme **Enz2**;  $[D > A]_{Enz2}$  describes the transfer of carbon from *D* to *A*, also catalyzed by **Enz2**; and  $[E > A]_{Enz2}$  describes the transfer of carbon from *E* to *A*, again catalyzed by **Enz2**. The three carbon atoms of *C* will form the first three carbons of *A* (in matrix notation a “1” indicates that a particular carbon atom is transferred from a reactant to a product, whereas a “0” indicates that a specific carbon atom is derived from a different reactant). *D* will contribute carbons 4 and 5 to *A*, and *E* will add the sixth carbon atom to the formation of *A* (Scheme 7). It is important to note that the AMMs do not contain information regarding the position of isotope label (which is specified by the label distribution vector); they just describe the transfer of carbon atoms from reactants to products.

**Scheme 7.** Example of generating atom mapping matrices.



As a third step, the metabolite labeling vector is multiplied by the appropriate AMM. By adding up the contributions of each reactant (product of AMM and the respective label distribution vector) modified by the corresponding reaction flux, we obtain the label distribution for the metabolic step specified in reaction (6) of Scheme 7:

$$A = [C > A]_{Enz2}C \cdot v_3 + [D > A]_{Enz2}D \cdot v_3 + [E > A]_{Enz2}E \cdot v_3$$

If all reactions using  $A$  as a reactant and all reactions forming  $A$  as a product are combined (reactions (1), (2), and (6) of Scheme 6), the following equation is obtained to describe label distribution:

$$b_I + v_2 \{ [C > A]_{Enz2}C + [D > A]_{Enz2}D + [E > A]_{Enz2}E \} = v_I \cdot A$$

The steady-state isotope distribution for the entire metabolic network under consideration (Scheme 4) can be calculated from the simultaneous solution of the above equation and those specifying label distributions for  $B$  and  $C$ . Using an iterative computational

approach, unknown metabolic fluxes can be estimated and the equations describing isotope distributions solved. The calculated values are then compared with experimentally determined isotope enrichment data and the computational process is repeated until satisfactory convergence is reached, while still meeting all stoichiometric constraints (Zupke and Stephanopoulos, 1994; Stephanopoulos et al., 1998).

### 2.2.3.2. Isotopomer balancing

A further development of the AMM concept involves the use of isotopomer mapping matrices (IMMs) (Jeffrey et al., 1991; Künneke et al., 1993; Schmidt et al., 1997), in which the expression of all isotopomer mass balances of a metabolite pool is achieved in a single matrix equation. In analogy to label distribution vectors of the AMM approach, IMMs use isotopomer distribution vectors (IDVs). Its notation is binary (hence the subscript ‘bin’) using a “0” for absence and a “1” for presence of isotope label. For metabolite *A* in the metabolic network depicted in Scheme 4, the IDV  $I_A$  will contain  $2^6 = 64$  elements ( $2^n$  isotopomers, where *n* is the number of carbons in the metabolite). The individual matrices shown in Scheme 7 are combined into one large matrix (Scheme 8)

**Scheme 8.** Example for generating isotopomer distribution vectors.

$$I_A = \begin{bmatrix} I_A(1) \\ I_A(2) \\ I_A(3) \\ I_A(4) \\ \dots \\ I_A(64) \end{bmatrix} = \begin{bmatrix} I_A(000000_{bin}) \\ I_A(000001_{bin}) \\ I_A(000010_{bin}) \\ I_A(000011_{bin}) \\ \dots \\ I_A(111111_{bin}) \end{bmatrix}$$

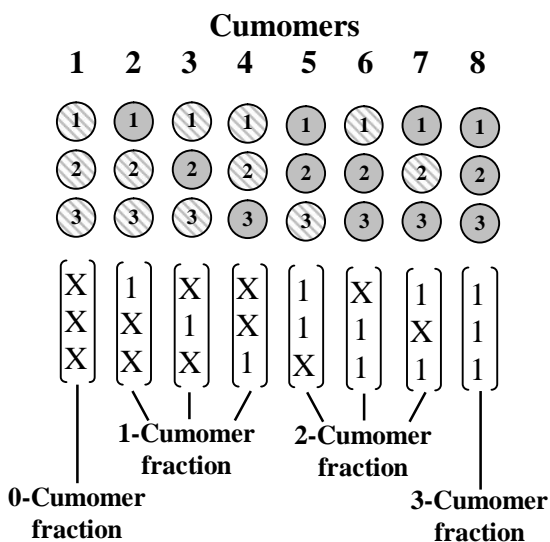
The IDVs of product molecules are obtained by matrix multiplication of IMMs (which inform about the reaction mechanism) and IDVs of the reactants (Schmidt et. al., 1997). IMMs contain all pairs of reactant isotopomers responsible for the synthesis of specific product isotopomers in all positions of the product IDV. Thus, IMMs can be used to identify the substrate isotopomer from which a specific product isotopomer was synthesized. There will be as many IMMs as the number of pairs of reactants and product molecules in a biochemical reaction (Schmidt et. al., 1997). IMMs can be generated from AMMs in an iterative process, which we illustrate for reaction (6) of Scheme 6 in Supplementary Material 1. For a detailed explanation of the analytical solution and the interpretation of these results, see Schmidt et al. (1999).

### 2.2.3.3. Cumomer balancing

Wiechert and colleagues (1999) developed an approach that simplifies the computational solution of balance equations where isotopomer variables are transformed into cumomer (*cumulated isotopomer*) variables. Cumomers are “virtual molecules” defining a set of isotopomers. The notation used for cumomers uses a “1” for a labeled carbon atom and an “X” for an atom that is either labeled or unlabeled, which means that the binary notation of isotopomer balancing (e.g.,  $C_{100\text{bin}}$ ) becomes a positional notation (e.g.,  $C_{1XX}$ ). Cumomers can be combined into fractions (0-cumomer fraction for  $C_{XXX}$ ; 1-cumomer fraction for  $C_{1XX}$ ,  $C_{X1X}$ ,  $C_{XX1}$ ; 2-cumomer fraction for  $C_{11X}$ ,  $C_{X11}$ ,  $C_{1X1}$ ; 3-cumomer fraction for  $C_{111}$ ). In analogy to the example of the three-carbon metabolite isotopomers of Scheme 4, the cumomer fractions are depicted as follows (gray shading is

used to indicate an isotope-labeled carbon, whereas gray and white stripes are used to indicate a carbon atom that is either isotope-labeled or not) (Scheme 9):

**Scheme 9.** Example showing all cumomers of a three-carbon metabolite.



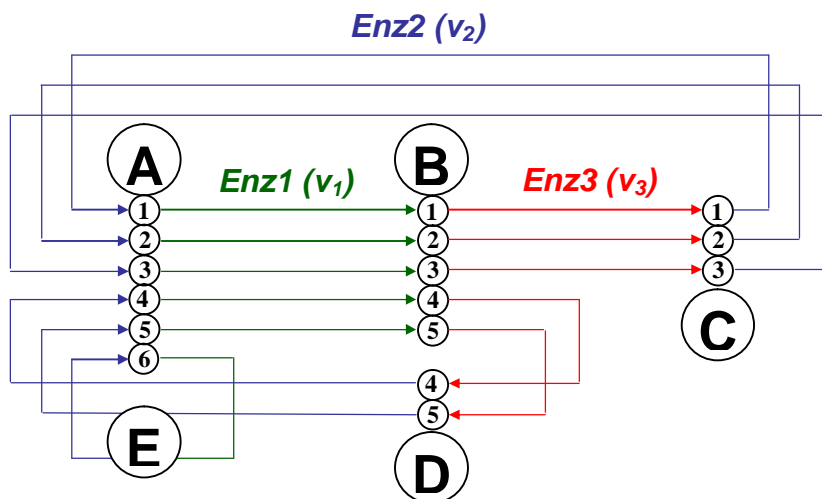
A novel framework for estimating flux distribution has been proposed (the elementary metabolite units (EMU) concept), which supposedly further simplifies the solution of isotopomer and cumomer equations; however, at the time of submission of this article only an abstract of the work was available online (Antoniewicz et al., 2007).

#### 2.2.3.4. Bondomer balancing

The bondomer balancing concept includes, in addition to isotopomer distributions, information about carbon-carbon connectivity in metabolites. Bondomers of a metabolite differ in the number and positions of C-C bonds that remain intact after an isotope-labeled substrate enters a metabolic network (van Winden et al., 2002). An advantage of the bondomer balancing method over isotopomer balancing is that the

number of required balance equations is lower, which makes bondomer balancing less computationally intense (van Winden et al., 2002). The bondomer balancing approach has been limited to experiments with [U- $^{13}\text{C}$ ]-labeled substrates ( $^2\text{H}$  or  $^{15}\text{N}$  do not provide information about the integrity of C-C bonds). An expansion of this concept to include C-H or C-N bonds is conceivable but would require [U- $^{13}\text{C}$ ; U- $^2\text{H}$ ] or [U- $^{13}\text{C}$ ; U- $^{15}\text{N}$ ]-labeled substrates and would be computationally challenging. Bondomer distributions can be constructed in the same way as isotopomer distributions, except that the binary digits representing labeled or unlabeled states in isotopomer balancing represent the origin and integrity of C-C bonds in bondomer balancing. Scheme 10 illustrates the concept of evaluating bond integrity using the metabolic network example of Scheme 4.

**Scheme 10.** Example of a bond-by-bond representation of the biochemical reaction network depicted in Scheme 4.





In the reactions shown in Scheme 10 several bonds are broken and the atoms of the metabolites are regrouped in every cycle until the metabolic network reaches isotopic steady-state. Using two-dimensional  $^{13}\text{C}$ -NMR, the proportion of intact  $^{13}\text{C}$ - $^{13}\text{C}$  bonds in a particular metabolite can be determined. In analogy to isotopomer balancing, bondomer balancing uses the terms C-C bond mapping matrices (CCMMs), C-C bond vectors (CCVs), bondomer distribution vectors (BDVs) and bondomer mapping matrices (BMMs), which replace AMMs, label distribution vectors, IDVs and IMM, respectively (van Winden et al., 2002). A new arithmetic approach for simulating bondomer distributions is based upon Boolean function mapping (Sriram and Shanks, 2004; Sriram et al., 2004). Using this method the topology of hypothetical pathways can be adjusted iteratively to match experimental isotopomer labeling results, thus allowing experimenters to test different pathway hypotheses.

#### **2.2.4. Metabolic flux analysis for modeling of plant metabolic networks**

Before attempting to conduct labeling experiments with subsequent mathematical modeling, it is important to understand to what extent the experimental design determines the possible outcomes. In principle, there are two types of isotope labeling approaches: whereas the transient approach requires sampling the time course as an isotopic label moves from an initial labeled substrate through the entire metabolic network, the steady-state approach is based upon a single measurement of labeled metabolic intermediates or end products at metabolic and isotopic steady-state. For metabolic networks that contain numerous reversible steps, futile cycles and/or enzymatic steps in different subcellular compartments (e.g., central carbon metabolism), the dynamic labeling approach is usually

not well suited to investigate flux distributions and the steady-state approach should be employed. However, for networks of smaller size and reduced complexity the dynamic labeling approach can be very powerful (Matsuda et al., 2003; Boatright et al., 2004; Roessner-Tunali et al., 2004). It should be noted that the methods listed under sections 2.3.1. to 2.3.4. apply only to steady-state modeling. For a more detailed discussion of the different isotope labeling approaches the reader is referred to an excellent recent review by Ratcliffe and Shachar-Hill (2006). In the upcoming paragraphs we are going to discuss examples from the literature to illustrate the complications when working with plant systems and highlight the success stories.

### **Steady-state labeling**

Dieuaide-Noubhani et al. (1995) used [1-<sup>14</sup>C]glucose as a substrate and the evolution of <sup>14</sup>CO<sub>2</sub> (the production of CO<sub>2</sub> at steady-state is constant) as an end point measurement to determine when metabolic and isotopic steady-states were reached. Since under regular conditions CO<sub>2</sub> formation increased continuously, the authors “prestarved” the root tips by incubating them in a medium that leads to a depletion of starch pools. Using this protocol a steady-state was reached after 10 h. Although this experimental trick rendered steady-state flux determinations feasible, it has to be questioned if such drastic treatments might displace the metabolic pathways studied by the authors from their usual steady-state. Thus, the fluxes determined by the authors might reflect metabolism under stress and could be quite different from those occurring in unstressed roots. Keeping this caveat in mind, the authors still provide valuable insights into the pathways utilized for the biosynthesis of particular cellular metabolite

pools. Labeling with [1-<sup>14</sup>C], [2-<sup>14</sup>C] or [6-<sup>14</sup>C]glucose revealed the importance of a plastid-localized pentose phosphate pathway for the formation of phospholipids and starch. Based on results obtained with feeding [1-<sup>13</sup>C]glucose and analysis of ethanol extracts by <sup>13</sup>C-NMR, it was concluded that cycling between hexose phosphates and triose phosphates occurred (resulting in randomization of label) and that a high turnover rate of sucrose (synthesis and degradation) in the cytosol was detectable. Further developments and expansions of this approach were later employed by the same group to investigate numerous central metabolic pathways during the growth cycle of tomato cell suspension cultures (Rontein et al., 2002). Roughly 30 fluxes were determined and allowed the authors to distinguish between pathways that appeared to be relatively resistant to changes in flux (e.g., glycolysis, pentose phosphate pathway, and the citric acid cycle) from those that reacted flexibly to flux alterations (e.g., various anabolic pathways). Edwards et al. (1998) extended these studies by measuring the <sup>13</sup>C fractional enrichment, using <sup>13</sup>C-NMR and GC-MS, in glucose 6-phosphate, L-alanine, L-glutamate, L-aspartate and malate when [1-<sup>13</sup>C]glucose was proffered to maize root tips grown. A simple precursor-product model allowed the authors to calculate the contributions of malic enzyme (EC 1.1.1.37), pyruvate kinase (EC 2.7.1.40) and phosphoenolpyruvate carboxylase (EC 4.1.1.31) to mitochondrial respiration. Fernie et al. (2001) studied heterotrophically grown tobacco callus cultures overexpressing a mammalian gene encoding 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (EC 2.7.1.105 / EC 3.1.3.46). This enzyme synthesizes fructose 2,6-bisphosphate, an allosteric regulator of the glycolytic enzyme fructose 1,6-bisphosphatase (EC 3.1.3.11) in photosynthetic tissues. Based on labeling experiments with [1-<sup>13</sup>C]glucose and

subsequent analysis of ethanol-soluble metabolite fractions by  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR, increases in the rate of cycling of triose phosphates to hexose phosphates were observed in transgenic lines (correlating with fructose-2,6-bisphosphate levels), whereas the rate of sucrose cycling was not affected. Unexpectedly, metabolic conversions of [U- $^{14}\text{C}$ ]glucose or [U- $^{14}\text{C}$ ]glycerol into organic acids, amino acids and lipids in transgenic lines with increased fructose 2,6-bisphosphate levels, when compared to appropriate controls, were very similar (Ferne et al., 2001). Eisenreich and coworkers developed a  $^{13}\text{C}$ -NMR-based approach to evaluate labeling patterns following steady-state labeling with a combination of both isotope-labeled and unlabeled substrates; the results can be used to infer the biosynthetic origin of intermediates in central carbon metabolism (Glawischnig et al., 2001, 2002) and, thanks to the development of an improved computational analysis, has provided information about glucose recycling in tobacco plants (Ettenhuber et al., 2005a) and has provided evidence that, based on the analysis of kernels from genetically diverse maize lines, fluxes through central carbon metabolism are quite robust (Ettenhuber et al., 2005b, Spielbauer et al., 2006). Some of the most informative isotope labeling-based metabolic flux studies with plants were those aimed at elucidating the control of central carbon metabolism in embryos of oilseed rape (Schwender and Ohlrogge, 2002; Schwender et al., 2003, 2004, 2006), which have been discussed in detail in a recent review (Ratcliffe and Shachar-Hill, 2006).

### **Dynamic labeling**

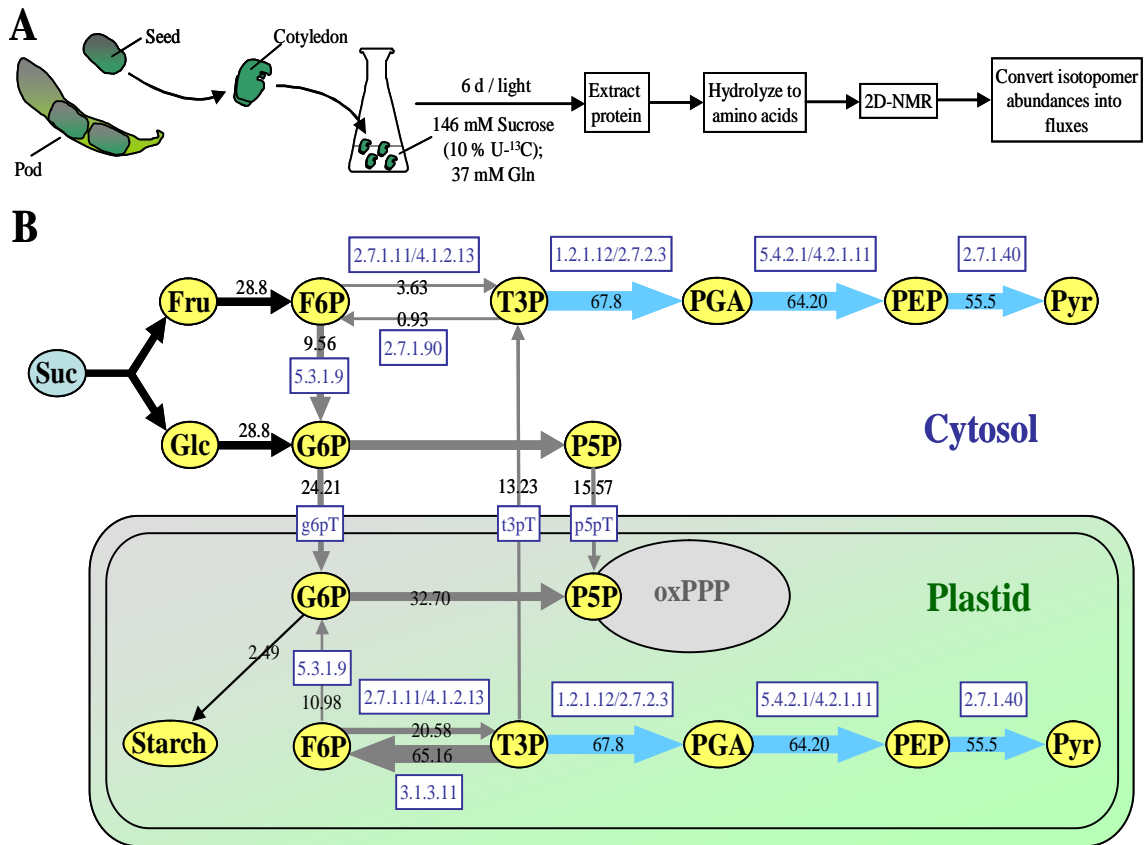
The dynamic labeling approach was utilized to characterize changes in phenylpropanoid metabolism in potato tubers after wounding (Matsuda et al., 2003). The

authors determined that the fluxes leading to *N-p*-coumaroyloctopamine (pCO) and chlorogenic acid (CGA), which are derived independently from a common precursor (*p*-coumaroyl-CoA), are not evenly distributed (4 : 1 flux ratio in favor of the pCO over the CGA branch). It was also shown that the breakdown rate of CGA was lower than that of pCO; because of its high turnover, pCO did not accumulate to high levels. The same approach was used in a follow-up paper studying the effect of oligosaccharide elicitor treatments on metabolic fluxes in potato tubers, which lead to a vastly increased flux into pCO (8.9-fold higher than in controls) and a decreased flux into CGA (2.7-fold lower than in controls) and metabolites derived from it (Matsuda et al., 2005). Despite the high pCO turnover rate, the elicitation resulted in a 25-fold transient increase in pCO levels. Boatright et al. (2004) studied a complex network of reactions metabolizing [<sup>2</sup>H<sub>5</sub>]phenylalanine to phenylpropanoids and benzenoids in excised petunia petals using the dynamic labeling approach. MFA revealed quantitative information about the branching of flux via CoA-dependent ( $\beta$ -oxidative) and CoA-independent (non- $\beta$ -oxidative) pathways. A model prediction associated the benzylbenzoate (BB), the role of which had not been previously defined, with the biosynthesis of benzoic acid and derived metabolites. The hypothesis that BB is a key intermediate in this pathway was substantiated by identifying and characterizing a petunia enzyme that catalyzes the formation of BB from benzoyl-CoA and benzyl alcohol. McNeil et al. (2000a) used mathematical modeling of [<sup>14</sup>C] and [<sup>33</sup>P] labeling data to determine the main route through the metabolic grid of choline (Cho) biosynthesis from ethanolamine in tobacco leaf disks. Follow-up work by the same group (McNeill et al., 2000b) assessed the role of compartmented fluxes (cytosolic, plastidial, vacuolar and apoplastic reaction were

considered) in determining the fate of Cho with respect to its conversion into glycine betaine (in plastids). Based on computer simulations several constraints for glycine betaine accumulation were suggested: (1) small cytosolic Cho pool, (2) low capacity for phospho-Cho synthesis, (3) high Cho kinase activity and (4) low activity for Cho uptake into plastids. The first three constraints would limit precursor synthesis and all enzymes involved in this process had been characterized, whereas the fourth constraint indicated that an as yet unidentified transport activity might be important in the regulation of glycine betaine biosynthesis. This study provides an example of how experimentally testable hypotheses can be derived from predictions generated by mathematical modeling. The redistribution of radiolabel (from [U-<sup>14</sup>C]glucose) and isotopic label (from [1-<sup>13</sup>C]glucose) in potato tubers subjected to low temperatures was investigated by Malone et al. (2006). It was concluded that the increased sugar accumulation during cold storage, which renders tubers unsuitable for further processing, was not caused by the lability of certain enzymes as proposed in earlier works. Alonso et al. (2005) used pulse-labeling with [<sup>14</sup>C]glucose to assess unidirectional rates of synthesis of major storage molecules (sucrose, starch and cell wall carbohydrates) and combined it with steady-state labeling using [1-<sup>13</sup>C] and [U-<sup>13</sup>C]glucose to quantify the extent of futile cycling in maize root tips. These studies identified a new glucose 6-phosphate to glucose cycle that is responsible for consuming a vast amount of ATP (roughly 40 % of the total ATP). Enzyme assays confirmed the presence of substantial glucose 6-phosphate phosphatase activity in root tips.

### **Highlight: central carbon metabolism in soybean embryos**

In the present review we are highlighting the development of an improved computational framework for the analysis of isotopomer abundances and its application to metabolic flux quantification in developing soybean embryos (Sriram et al., 2004). Labeling experiments of *in vitro* cultured, excised cotyledons were performed with a mixture of [U-<sup>13</sup>C]glucose, glucose with natural <sup>13</sup>C abundance, and glutamine (Figure 2.2). Storage protein and starch were extracted, hydrolyzed, and the resulting amino acids, sugars and derived hydrolyzates were subjected to analysis by two-dimensional [<sup>1</sup>H, <sup>13</sup>C] NMR, primarily using the heteronuclear single quantum correlation (HSQC) pulse sequence. Assessment of cross peak intensities in NMR spectra and the evaluation of <sup>13</sup>C-<sup>13</sup>C scalar coupling allowed isotopomeric abundances to be quantified for the aliphatic carbon atoms of 16 amino acids, the aromatic carbon atoms of phenylalanine, tyrosine and histidine, and for hexose sugars (which were detected as their hydrolysis products levulinic acid and hydroxyacetone) without prior purification of the analytes. Based on the available biochemical information in the literature, the authors reconstructed a model of the metabolic network involved in the synthesis of sink metabolites (amino acids that are integrated into storage proteins and sugar phosphates that are converted into transitory starch). If the pathway that is involved in converting a precursor into a sink end product was known, the NMR signal intensities measured for sink metabolites could be used to infer the expected isotopomeric composition of this precursor.



**Figure 2.2.** Metabolic flux analysis of central carbon metabolism in developing soybean embryos (based upon data from Sriram et al., 2004). **A**, Flowchart of pod harvest, embryo sampling, isotope feeding, protein extraction and processing, and data analysis; **B**, Visualization of the control of glycolysis using the BioPathAt tool (Lange and Ghassemian, 2005). The widths of reaction arrows correlate with net fluxes. Net flux values are also given as  $\mu\text{mol} (\text{day} \cdot \text{cotyledon})^{-1}$ . Enzymes are represented by their EC numbers. Black arrows indicate fluxes that are restricted to a particular subcellular compartment, gray arrows depict fluxes that occur in both the cytosol and plastids but can be distinguished, and light blue arrows indicate steps occurring in both compartments that can not be distinguished based upon the data in Sriram et al. (2004). The values given in this figure are different from those published in the original paper; they are based upon a correction published by Sriram et al. (Plant Physiology (2006) 142, 1771).



By integrating the mathematical concepts of isotopomer balancing (described in sections 2.3.2. to 2.3.4. of this review and references cited therein) with global optimization methods, Sriram et al. (2004) developed a computational tool, termed NMR2Flux, to convert isotopomer abundances into fluxes in an automated fashion. Figure 2 summarizes these flux determinations for several steps of glycolysis (which is just an extract of the entire network considered by the authors). Forward and backward fluxes were calculated for steps depicted with gray reaction arrows in Figure 2.2, although only the direction and values of total net fluxes are indicated. Sriram et al. (2004) were also able to distinguish between parallel fluxes in different compartments generated by isozyme activities (examples in Figure 2.2 would be steps associated with the same EC number in both the cytosol and plastids), whereas in other cases cytosolic and plastidial fluxes remained indistinguishable (light blue arrows in Figure 2.2). Taken together the study by Sriram et al. (2004) represents the single most comprehensive metabolic flux analysis performed in a plant system to date. This integrative approach, which could be extended to other crop plants, has potential for providing valuable data to inform metabolic engineering and/or molecular breeding efforts aimed at enhancing key metabolic traits.

### **2.3. Kinetic analysis of metabolic networks**

Stoichiometric network models can be generated using metabolic reconstructions that are based upon predictions of the coding capacity of sequenced genomes. This modeling approach is a first logical step in the exploration of the basic properties of the

metabolic network. As outlined in Supplementary Material 1, linear programming tools can be used to calculate optimal flux distributions in the network when physicochemical constraints and objective functions (e.g., the system is geared toward maximizing the yields of particular sink products) are applied. Methods of metabolic pathway analysis can then be used to obtain the feasible set of pathways operating in a given metabolic network. Isotopomer balancing can be employed to assign quantitative fluxes to specific branches of pathways or even individual steps. However, when scientists have aimed at changing fluxes by metabolic engineering (without disturbing the overall cell physiology), it has been very hard to determine the expression levels of which enzyme(s) needed to be modified. In complex networks (and almost all plant pathways have to be considered as being part of larger networks) intuitive approaches to the optimization of fluxes have remained unsatisfactory. Mathematical models that describe the dynamic properties of metabolic networks (predicting variation of metabolite concentrations over time) build on known rate equations (expressing enzyme properties and behavior in the presence of effectors) for individual enzymatic reactions and the definition of a metabolic network under consideration as a set of interdependent ODEs. Usually, algebraic solutions can not be obtained for these systems of ODEs but various software packages have been introduced that allow calculating numerical solutions (iterative process of approximation and error correction) (for an overview of available software see Table 2.1). Once a kinetic model has been constructed it is possible to calculate the relative impact of an enzymatic step on the concentration of a metabolite the production of which one wishes to optimize. Although potentially very powerful and information-rich, such modeling efforts suffer from the disadvantage that most kinetic data are obtained with

purified enzyme preparations in test tubes (*in vitro* kinetics), which might not be good reflections of kinetic characteristics within the cell (*in vivo* kinetics) (Steuer et al., 2006) and it has been suggested that approximative approaches to infer *in vivo* enzyme kinetics should be employed (Teusink et al., 2000; Heijnen, 2005).

### 2.3.1. Case study of kinetic modeling and dynamic simulation

We will consider the metabolic network described in Scheme 1 under the assumption that the transport fluxes  $\mathbf{b}_1$ ,  $\mathbf{b}_2$  and  $\mathbf{b}_3$  exert negligible control over flux; this means that only the irreversible reaction fluxes  $\nu_1$ ,  $\nu_2$  and  $\nu_3$  need to be considered. To obtain the mass balances for metabolites  $A$ ,  $B$  and  $C$ , the reaction network stoichiometry (I) is combined with Michaelis-Menten type enzyme kinetics (II) for each enzymatic step:

(I) Stoichiometric  
balances

$$\frac{dA}{dt} = -\nu_1 - \nu_2$$

$$\frac{dB}{dt} = \nu_1 - \nu_3$$

$$\frac{dC}{dt} = \nu_3 + \nu_2$$

(II) Michaelis-Menten rate  
constants

$$\nu_1 = v_{max(Enz1)} \cdot \frac{A}{A + K_m(Enz1)}$$

$$\nu_3 = v_{max(Enz3)} \cdot \frac{A}{A + K_m(Enz3)}$$

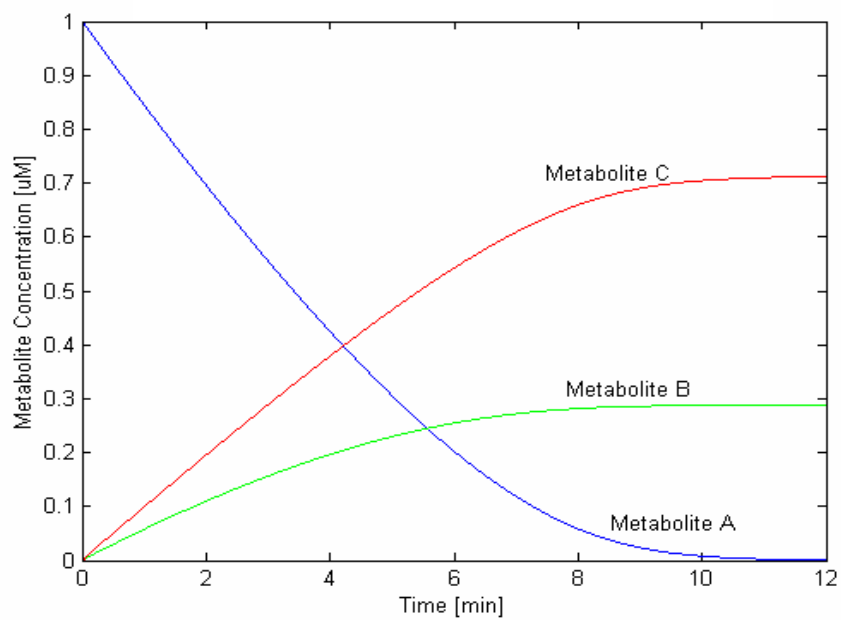
$$\nu_2 = v_{max(Enz2)} \cdot \frac{A}{A + K_m(Enz2)}$$

The variation of metabolite concentration over time ( $dM/dt$ ) is equal to the difference of its rate of formation and its rate of consumption. Let us assume that we have measured an initial substrate concentration ( $[A] = 1 \mu\text{M}$ ) and that we have obtained the following values for  $K_m$  and  $V_{max}$ :  $K_m(Enz1) = 0.8 \mu\text{M}$ ;  $K_m(Enz2) = 0.1 \mu\text{M}$ ;  $K_m(Enz3) = 0.004 \mu\text{M}$ ;  $V_{max(Enz1)} = 0.0018 \mu\text{M} \cdot \text{sec}^{-1}$ ;  $V_{max(Enz2)} = 0.0018 \mu\text{M} \cdot \text{sec}^{-1}$ ;  $V_{max(Enz3)} =$

0.000012  $\mu\text{M} \cdot \text{sec}^{-1}$ . The simultaneous solution of the system of ODEs (for our example we used the 'ode45' function in the MATLAB® software; the code is available in Supplementary Material 1) then allows for the dynamic simulation of the changes in metabolite levels (Figure 2.3).

### **2.3.2. Kinetic models of plant metabolic networks**

The majority of complex kinetic models generated to describe the control of plant metabolism are focused on photosynthesis and related pathways of central carbon metabolism (Table 2.2), which are among the most studied pathways in plants. Morgan and Rhodes (2002) have discussed the outcomes of these modeling efforts in an excellent recent review article. In this chapter we will thus not review these papers again but have decided to make a few comments about future directions. Most pathways of primary metabolism are somewhat conserved among eukaryotes and, if kinetic constants should not be available for a particular plant enzyme, they can be estimated based on the values obtained with the same enzyme from another organism. There is, of course, a danger when kinetic constants are inferred from the literature; thus, kinetic models need to be evaluated against experimental data. In this regard, measurements of metabolite pool sizes are of critical importance to test the performance of mathematical models. However, the quantification of pathway intermediates is analytically challenging and often neglected. Metabolite profiling efforts will need to emphasize increasing the spectrum of relevant metabolic intermediates (which usually do not accumulate at high levels and might be chemically unstable), so that an approximation of modeling and experimental results can be performed.



**Figure 2.3.** Dynamic simulation of changes in metabolite concentrations over time for the metabolic network depicted in Scheme 1 (for assumed kinetic parameters and other details see text).

### 2.3.3. Metabolic control analysis

Metabolic control analysis (MCA) is a method for analyzing how control over fluxes and metabolite concentrations is distributed among the enzymes that constitute a metabolic pathway (or a network of pathways). Flux control is determined by applying experimental perturbations (one at a time) of each enzymatic step of interest (e.g., by over-expressing a particular enzyme in transgenic plants) and measuring the effect on the variable of interest after the system has reached a new steady-state. The magnitude of change in a network variable as a response to an experimental perturbation is expressed as a control coefficient such as the flux control coefficient (FCC):

$$C_{E_i}^J = \frac{\partial J / J}{\partial E_i / E_i} = \frac{\partial \ln J}{\partial \ln E_i}$$

where  $C_{E_i}^J$  is the FCC of the  $i^{\text{th}}$  enzyme,  $J$  is the steady-state flux, and  $E_i$  is the specific activity of  $i^{\text{th}}$  enzyme.

According to the summation theorem, the sum of all FCCs in a metabolic network is equal to unity (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Giersch, 1988; Reder, 1988):

$$\sum_{i=1}^n C_{E_i}^J = 1 \quad \text{Summation theorem}$$

where  $n$  corresponds to the number of enzymes involved in the metabolic system.

An important consequence of the summation theorem is that FCCs reflect global properties of a metabolic network; since control of flux is shared by all enzymes, the existence of rate-limiting steps is very unlikely. How does one actually assign FCCs? *In vitro* kinetic data that have been collected for the enzymes of a particular pathway can be used to calculate elasticity coefficients, which in turn can be used to calculate FCCs. Elasticity coefficients (ECs) establish a relationship between enzyme activity and its potential to control flux. The effect of a metabolite ( $M$ ) on enzyme activity is quantified as the fractional change in the reaction rate ( $v$ ) due to an infinitesimal change in the concentration of  $M$  (Fell, 1997):

$$\mathcal{E}_M^{Ei} = \frac{\partial v_i / v_i}{\partial M / M} = \frac{\partial \ln v_i}{\partial \ln M}$$

If  $M$  stimulates the rate of reaction (activator),  $\mathcal{E}_M^{Ei}$  will be greater than zero; if  $M$  slows the reaction down (inhibitor),  $\mathcal{E}_M^{Ei}$  will be negative. ECs are not systemic properties (unlike FCCs), but rather indicate how sensitive individual enzymes are to perturbations. The connectivity theorem of MCA states that the sum of the products of the FCCs and ECs of all ( $i$ ) enzymatic steps affected by  $M$  equals zero:

$$\sum_{i=1}^n C_{Ei}^J \mathcal{E}_M^{Ei} = 0 \quad \text{Connectivity theorem}$$

where  $n$  is the total number of enzymes in the system.

In analogy to FCCs, MCA also defines coefficients that account for the effect of parameters such as enzyme activity on intracellular metabolite concentration; these are

termed concentration control coefficients (CCCs). These coefficients specify the relative change in the level of the  $j^{\text{th}}$  metabolite when the activity of the  $i^{\text{th}}$  enzyme is changed (note that CCCs can have both positive and negative values):

$$C_{Ei}^{Mj} = \frac{E_i}{M_j} \cdot \frac{\partial M_j}{\partial E_i} = \frac{\partial \ln M_j}{\partial \ln E_i}$$

The summation theorem also applies and states that the sum of all CCCs equals zero:

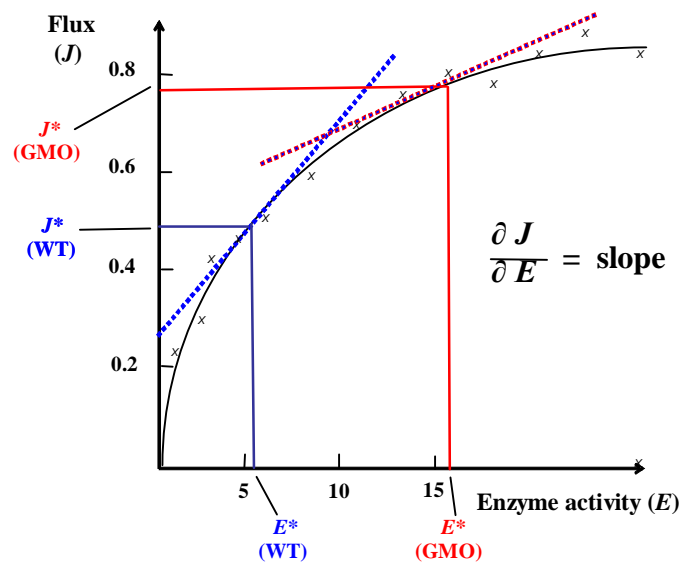
$$\sum_{i=1}^n C_{Ei}^{Mj} = 0$$

MCA thus relates the kinetic properties of individual enzymes (local attributes) to the properties of the entire pathway or network of pathways (global attributes).

### 2.3.3.1. Case study of metabolic control analysis

Two different approaches can be used to determine FCCs. The indirect approach requires the determination of elasticities which, based upon the connectivity theorem, can be used to calculate FCCs. The direct experimental determination of FCCs involves the variation of enzyme activity and/or kinetic parameters (e.g., by overexpressing an enzyme using transgenic approaches or inhibitor titration) and quantifying changes under steady-state conditions (Delgado and Liao, 1992) (Figure 2.4). We will again use the example shown in Scheme 1 to illustrate the use of the direct approach. Let us assume that we have generated transgenic plants with varying levels of expression of the gene encoding *Enz2* (and thus varying levels of *Enz2* enzyme activity).





**Figure 2.4.** Determination of flux control coefficients using the direct method of metabolic control analysis (for details see text). Abbreviations: GMO, transgenic plant; WT, wild-type plant.

By plotting enzyme activity versus flux as in Figure 2.4, a hyperbolic curve is obtained (each 'x' symbolizes an enzyme activity/flux pair for one plant line). The graph represents the steady state pathway flux  $J$  as a function of the activity of an enzyme in the metabolic network under consideration. The calculation of the FCC for an enzyme at a specific enzyme activity ( $E^*$ ), is performed by taking the slope of the tangent at that enzyme activity  $(dJ/dE)|_{E=E^*}$  and multiplying it by a scaling factor. The scaling factor includes the enzyme activity normalized with respect to the steady state flux  $J^*$ :  $FCC = (dJ/dE)|_{E=E^*} \cdot E/J = (dJ/J)/(dE/E)$ .

The scaling process makes the FCCs dimensionless, reflecting rates of change in proportion to the enzyme activity and steady state flux, at the fixed point ( $E^*$ ,  $J^*$ ). In our example (Figure 2.4),  $E^*$  for the wild-type plant is 5 (arbitrary units) and for the transgenic plant 16;  $J^*$  for wild-type plant is 0.5 (arbitrary units) and for the transgenic plant 0.75. The slope for the wild-type plant is 0.033 and for the transgenic plant 0.012. Thus, the FCCs for *Enz2* in the wild-type plants can be calculated as  $0.033 \cdot 6/0.45 = 0.44$  and for the transgenic plant we obtain  $0.012 \cdot 16/0.75 = 0.26$ . Note that the FCC for *Enz2* in the transgenic plant overexpressing the gene encoding *Enz2* is actually lower than that for *Enz2* in the untransformed control plants, indicating that a 'bottleneck' created by *Enz2* in the control plant has been relieved in the transgenic plant. At the same time, the FCCs for all other enzymes in the transgenic plant will have changed as well and another enzyme might now be responsible for a 'bottleneck'. This example illustrates that an enzyme that might have been regarded as being rate-limiting using classical biochemical terms does not exert appreciable control over the pathway in the transgenic plant. Flux control is a dynamic process and all enzymes of a metabolic

network contribute to it at varying degrees, which can change depending on the experimental conditions.

### **2.3.3.2. Metabolic control analysis for modeling plant metabolic networks**

#### **Indirect metabolic control analysis**

Poolman et al. (2000) introduced an approach to modeling photosynthetic pathways (in particular Calvin cycle, starch metabolism, and triose phosphate export) that distinguishes between two different steady-states depending on the magnitude of the carbon assimilatory flux. Based on an indirect MCA with regard to Calvin cycle enzymes, only sedoheptulose biphosphatase (EC 3.1.3.37) and ribulose biphosphate carboxylase/oxidase (Rubisco; EC 4.1.1.39) exert any significant control over assimilation. Once carbon is assimilated, other enzymes, most notably the triose phosphate transporter, become important in controlling the metabolic fate of assimilates. Thus, a simple strategy to increase carbon assimilation by transgenic over-expression of a small number of genes involved in the Calvin cycle is unlikely to be successful. Daae et al. (1999) used MCA to evaluate the feasibility of producing a polyhydroxyalkanoate polymer in transgenic plants. This analysis indicated that the adjustment of the ratio of the two precursor monomers would depend on growth conditions (light or dark), the expression levels of the four transgenes, and the availability of enzyme cofactors. For the development of a commercial production system, all of these variables would have to be carefully considered and controlled. Curien et al. (2003) developed a kinetic model around phosphohomoserine, an intermediate metabolite at the branchpoint of methionine

and threonine biosynthesis. The authors also considered that, based on *in vitro* assays, flux is regulated by various allosteric controls. An iterative approach of mathematical modeling and *in vitro* experimentation was then employed to develop a model that can be used to interrogate which changes in methionine and threonine levels are to be expected when the amounts of specific components of the metabolic network are modulated. The role of redistribution of flux control in the competing branches of glutathione (GSH) and phytochelatin metabolism under Cadmium stress conditions was studied by Mendoza-Cózatl and Moreno-Sánchez (2006). In unstressed plants GSH levels are controlled by demand, whereas under Cadmium stress various enzymes control the accumulation levels of GSH and phytochelatins. Transgenic approaches to increase phytochelatin amounts (and thus Cadmium tolerance) without affecting GSH levels would need to take into account that a balance between competing pathways needs to be maintained.

### **Direct metabolic control analysis**

Kruckeberg et al. (1989) evaluated the significance of plastidial and cytosolic isoforms of phosphoglucose isomerase (EC 5.3.1.9) in controlling the synthesis of starch and sucrose using a direct MCA approach based on results obtained with decreased-activity mutants. The plastidial isozyme was found to exert considerable control over starch biosynthesis only under high light and CO<sub>2</sub> conditions, whereas the cytosolic isozyme exerted appreciable control only in the dark. The primary role of the cytosolic isozyme appeared to be its involvement in sucrose synthesis. Direct MCA to estimate control coefficients for an individual enzyme in transgenic plants with varying expression (and thus specific activity) levels was also performed for Rubisco in tobacco plants

(Quick et al., 1991; Stitt et al., 1991), the glycolysis enzyme phosphofructokinase (EC 2.7.1.11) in potato tubers (Thomas et al., 1997a,b), the gluconeogenic enzyme isocitrate lyase (EC 4.1.3.1) in castor bean seedling endosperm (Runquist and Kruger, 1999), the carotenogenic enzyme phytoene synthase (EC 2.5.1.32) in tomato fruits (Fraser et al., 2002), and the lipid biosynthetic enzyme diacylglycerol acyltransferase (EC 2.3.1.20) in olive and oil palm tissue cultures (Ramli et al., 2005). Inhibitor titration studies allowed direct MCA to be employed for evaluating the importance of acetyl coenzyme A carboxylase (EC 6.4.1.2) in regulating lipid biosynthesis (Page et al., 1994), for determining the control of ammonia assimilation in isolated barley chloroplasts (Baron et al., 1994), and for investigating respiratory flux in isolated turnip mitochondria (Padovan et al., 1989). The outcomes of these direct MCA studies have been covered in previous reviews (Morgan and Rhodes, 2002; Fernie et al., 2005) and will not be discussed in detail here.

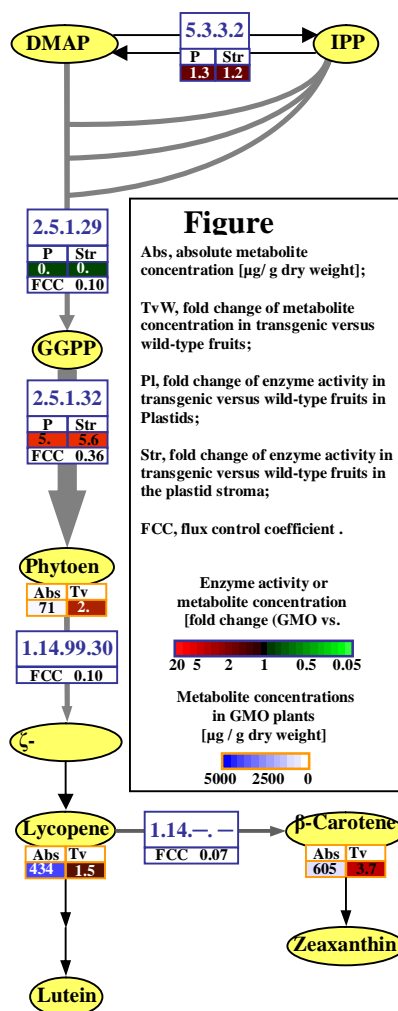
### **Top down control analysis**

If kinetic information is not available for all enzymes in a metabolic pathway (or a more complex network of pathways), the application of 'top down' control analysis (TDCA), an MCA approach that groups several reactions in 'black boxes' around key intermediates, may provide useful information (Brown et al., 1990). TDCA was successfully employed to examine the relative importance of two blocks of lipid biosynthesis (plastidial fatty acid biosynthesis and lipid assembly in the cytosol) in olive and oil palm callus cultures (Ramli et al., 2002). Although the biosynthesis block exerted higher control than the lipid assembly block, both blocks exert significant control; thus,

an over-expression of a single enzyme (or even a small set of enzymes) in transgenic olive or oil palm cultures is unlikely to result in increased oil yield. In general, all of the above-mentioned MCA studies have confirmed that the presence of a true rate-limiting enzyme is an extremely rare case.

### **Highlight: carotenoid biosynthesis in tomato fruits**

We will discuss one recent paper in more detail as an example to illustrate this fact (Fraser et al., 2002). The roots of this project go back to the early 1990s, when attempts to increase carotenoid content in transgenic tomato plants by over-expressing a constitutive copy of the carotenoid biosynthetic enzyme phytoene synthase (PSY; EC 2.5.1.32) were frustrated by ectopic pigment production and a stunted growth phenotype (Fray and Grierson, 1993). Follow-up work established that dwarfism was caused by a lack of precursor availability for the biosynthesis of gibberellins (Fray et al., 1995), which, like carotenoids, are derived from geranylgeranyl diphosphate as an intermediate (Figure 2.5). Fraser et al. (2002) employed constructs featuring a fruit-specific promoter and a transit sequence, thus targeting recombinant PSY protein to chromoplasts of tomato fruits. The FCC of PSY was determined to be 0.36, whereas FCCs for other enzymes considered were at 0.1 or lower (Figure 2.5). The 5 to 10-fold over-expression of PSY (at the level of specific enzyme activity) lead only to a roughly two-fold increase in total carotenoids. The FCC decreased from 0.36 (wild-type controls) to 0.15 (transgenic plants) with only very moderate increases in carotenoid content, indicating that flux control is shared by several enzymes in the carotenoid pathway.



**Figure 2.5.** Metabolic control analysis of carotenoid biosynthesis in tomato fruits (based upon data from Fraser et al., 2002). The BioPathAt tool was used to visualize (color-coded) ratios of enzyme activities and metabolite concentrations (transgenic versus wild-type plants) and absolute metabolite concentrations. Enzymes are represented by their EC numbers. The widths of reaction arrows correlate with flux control coefficients. Abbreviations for metabolites: DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; IPP, isopentenyl diphosphate.

## **2.4. Conclusions**

Various innovative approaches for metabolic flux analysis have been employed with plants in recent years (e.g., elementary modes analysis and isotopomer balancing), with other methods having potential although they have not yet been utilized (e.g., extreme pathway analysis). Metabolic flux analysis is a very active research field and has yielded valuable insights particularly into the regulation of central carbon metabolism. In contrast, kinetic approaches to estimate flux control were pursued by numerous groups in the 1990s and early 2000s but only a few studies have been published in this area recently. This might in part be due to a shift in governmental funding priorities, which have led to vast increases in the number of genomics-based studies, whereas our knowledge regarding the kinetic properties of individual metabolic enzymes has not kept pace. Both approaches depend on each other to provide an in-depth understanding of flux control and it should thus be a high priority for funding agencies to help reinvigorate classical biochemistry.

### **Acknowledgements**

R.R.E would like to thank the Fulbright Program and the University of Antioquia for scholarships; B.M.L would like to thank the National Science Foundation (award 0520140) and the Agricultural Research Center at Washington State University for funding. The authors would also like to thank Dr. W. Wiechert (University of Siegen, Germany) for advice on the chapter describing the isotopomer approach to flux analysis. We would also like to thank two anonymous reviewers for valuable comments that have considerably improved this manuscript.



## 2.5 References

- Alves, R., Antunes, F., Salvador, A., 2006. Tools for kinetic modeling of biochemical networks. *Nature Biotechnology* 24, 667-672.
- Alonso, A.P., Vigeolas, H., Raymond, P., Rolin, D., Dieuaide-Noubhani, M., 2006. A new substrate cycle in plants. Evidence for a high glucose-phosphate-to-glucose turnover from in vivo steady-state and pulse-labeling experiments with [<sup>13</sup>C]glucose and [<sup>14</sup>C]glucose. *Plant Physiology* 138, 2220-2232.
- Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2007. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic Engineering* 9, 68-86.
- Baron, A.C., Tobin, T.H., Wallsgrave, R.M., Tobin, A.K., 1994. A metabolic control analysis of the glutamine synthetase/glutamate synthase cycle in isolated barley (*Hordeum vulgare* L.) chloroplasts. *Plant Physiology* 105, 415-424.
- Bell, S.L., Palsson, B.Ø., 2005. *ExpRA*: a program for calculating extreme pathways in biochemical reaction networks. *Bioinformatics* 21, 1739-1740.
- Blasius, B., Beck, F., Lüttge, U., 1997. A model for photosynthetic oscillations in Crassulacean acid metabolism (CAM). *Journal of Theoretical Biology* 184, 345-351.
- Blasius, B., Beck, F., Lüttge, U., 1998. Oscillatory model of Crassulacean acid metabolism: structural analysis and stability boundaries with a discrete hysteresis switch. *Plant, Cell and Environment* 21, 775-784.
- Blasius, B., Neff, R., Beck, F., Lüttge, U., 1999. Oscillatory model of Crassulacean acid metabolism with dynamic hysteresis switch. *Proceedings of the Royal Society of London B: Biological Sciences* 266, 93-101.
- Boatright, J., Negre, F., Chen, X., Kish, C.M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D., Dudareva, N., 2004. Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiology* 135, 1993-2011.
- Brown, G.C., Hafner, R.P., Brand, M.D., 1990. A 'top-down' approach to the determination of control coefficients in metabolic control theory. *European Journal of Biochemistry* 188, 321-325.
- Caiveau, O., Fortune, D., Cantrel, C., Zachowski, A., Moreau, F., 2001. Consequences of omega-6-oleate desaturase deficiency on lipid dynamics and functional properties of mitochondrial membranes of *Arabidopsis thaliana*. *Journal of Biological Chemistry* 276, 5788-5794.

Chen, D.X., Coughenour, M.B., Knapp, A.K., Owensby, C.E., 1994. Mathematical simulation of C<sub>4</sub> grass photosynthesis in ambient and elevated CO<sub>2</sub>. *Ecological Modelling* 73, 63-80.

Collatz, G.J., Ribas-Carbo, M., Berry, J.A., 1992. Coupled photosynthesis-stomatal conductance model for leaves of C<sub>4</sub> plants. *Australian Journal of Plant Physiology* 19, 519-538.

Curien, G., Ravanel, S., Dumas, R., 2003. A kinetic model of the branch-point between the methionine and threonine biosynthesis pathways in *Arabidopsis thaliana*. *European Journal of Biochemistry* 270, 4615-4627.

Dae, E.B., Dunnill, P., Mitsky, T.A., Padgett, S.R., Taylor, N.B., Valentin, H.E., Gruys, K.J., 1999. Metabolic modeling as a tool for evaluating polyhydroxyalkanoate copolymer production in plants. *Metabolic Engineering* 1, 243-254.

Delgado, J., Liao, J.C., 1992. Determination of Flux Control Coefficients from transient metabolite concentrations. *Biochemical Journal* 282, 919-927.

DellaPenna, D., 2001. Plant metabolic engineering. *Plant Physiology* 125, 160-163.

Dhar, P., Meng, T.C., Somani, S., Ye, L., Sairam, A., Chitre, M., Hao, Z., Sakharkar, K., 2004. Cellware - a multi-algorithmic software for computational systems biology. *Bioinformatics* 20, 1319-1321.

Dieuaide-Noubhani, M., Raffard, G., Canioni, P., Pradet, A., Raymond, P., 1995. Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from <sup>13</sup>C- or <sup>14</sup>C-labeled glucose. *Journal of Biological Chemistry* 270, 13147-13159.

Edwards, S., Nguyen, B.T., Do, B., Roberts, J.K.M., 1998. Contribution of malic enzyme, pyruvate kinase, phosphoenolpyruvate carboxylase, and the Krebs cycle to respiration and biosynthesis and to intracellular pH regulation during hypoxia in maize root tips observed by nuclear magnetic resonance imaging and gas chromatography-mass spectrometry. *Plant Physiology* 116, 1073-1081.

Ettenhuber, C., Radykewicz, T., Kofer, W., Koop, H.U., Bacher, A., Eisenreich, W., 2005a. Metabolic flux analysis in complex isotopolog space. Recycling of glucose in tobacco plants. *Phytochemistry* 66, 323-335.

Ettenhuber, C., Spielbauer, G., Margl, L., Hannah, L.C., Gierl, A., Bacher, A., Genschel, U., Eisenreich, W., 2005b. Changes in flux pattern of the central carbohydrate metabolism during kernel development in maize. *Phytochemistry* 66, 2632-2642.

Farquhar, G.D., von Caemmerer, C., Berry, J.A., 1980. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta* 149, 78-90.

- Fell, D.A., 1997. Understanding the control of metabolism. Portland Press, London, UK.
- Fernie, A.R., Roscher, A., Ratcliffe, R.G., Kruger, N.J., 2001. Fructose 2,6-bisphosphate activates pyrophosphate: fructose-6-phosphate 1-phosphotransferase and increases triose phosphate to hexose phosphate cycling in heterotrophic cells. *Planta* 212, 250-263.
- Fernie, A.R., Geigenberger, P., Stitt, M., 2005. Flux an important, but neglected, component of functional genomics. *Current Opinion in Plant Biology* 8, 174-182.
- Fraser, P.D., Romer, S., Shipton, C.A., Mills, P.B., Kiano, J.W., Misawa, N., Drake, R.G., Schuch, W., Bramley, P.M., 2002. Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. *Proceedings of the National Academies of Sciences of the United States of America* 99, 1092-1097.
- Fray, R.G., Grierson, D., 1993. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol. Biol.* 22, 589-602.
- Fray, R.G., Wallace, A., Fraser, P.D., Valero, D., Hedden, P., Bramley, P.M., Grierson, D., 1995. Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *The Plant Journal* 8, 693-701.
- Fridlyand, L.E., 1998. Independent changes of ATP/ADP or Delta-pH could cause oscillations in photosynthesis. *Journal of Theoretical Biology* 193, 739-741.
- Fridlyand, L.E., Backhausen, J.E., Scheibe, R., 1998. Flux control of the malate valve in leaf cells. *Archives of Biochemistry and Biophysics* 349, 290-298.
- Fridlyand, L.E., Scheibe, R., 1999. Regulation of the Calvin cycle for CO<sub>2</sub> fixation as an example for general control mechanisms in metabolic cycles. *Biosystems* 51, 79-93.
- Fridlyand, L.E., Scheibe, R., 2000. Regulation in metabolic systems under homeostatic flux control. *Archives of Biochemistry and Biophysics* 374, 198-206.
- Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H., 2003. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks, *Biosilico* 1, 159-162.
- Gagneur, J., Klamt, S., 2004. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* 5, 175-175.
- Giersch, C., 1988. Control analysis of biochemical pathways: a novel procedure for calculating control coefficients, and an additional theorem for branched pathways. *Journal of Theoretical Biology* 134, 451-462.

Giersch, C., 2000. Mathematical modelling of metabolism. *Current Opinion in Plant Biology* 3, 249-253.

Glawischnig, E., Gierl, A., Tomas, A., Bacher, A., Eisenreich, W., 2001. Retrobiosynthetic nuclear magnetic resonance analysis of amino acid biosynthesis and intermediary metabolism. Metabolic flux in developing maize kernels. *Plant Physiology* 125, 1178-1186.

Glawischnig, E., Gierl, A., Tomas, A., Bacher, A., Eisenreich, W., 2002. Starch biosynthesis and intermediary metabolism in maize kernels. Quantitative analysis of metabolite flux by nuclear magnetic resonance. *Plant Physiology* 130, 1717-1727.

Gross, L.J., Kirschbaum, M., Pearcy, R.W., 1991. A dynamic model of photosynthesis in varying light taking account of stomatal conductance, C[3]-cycle intermediates, photorespiration and Rubisco activation. *Plant, Cell and Environment* 14, 881-893.

He, D., Edwards, G.E., 1996. Estimation of diffusive resistance of bundle sheath cells to CO<sub>2</sub> from modeling of C<sub>4</sub> photosynthesis. *Photosynthesis Research* 49, 195-208.

Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European Journal of Biochemistry* 42, 89-95.

Heinrich, R., Schuster, S., 1996. *The regulation of cellular systems*. Chapman and Hall, New York.

Heijnen, J.J., 2005. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and Bioengineering* 91, 534-545.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI - a COMplex PATHway SIMulator. *Bioinformatics*. 22, 3067-3074.

Jeffrey, F.M.H., Rajagopal, A., Malloy, C.R., Sherry, A.D., 1991. <sup>13</sup>C-NMR: a simple yet comprehensive method for analysis of intermediary metabolism. *Trends in Biochemical Sciences* 16, 5-10.

Kacser, H., Burns, J.A., 1973. The control of flux. *Symposium of the Society of Experimental Biology* 27, 65-104.

Kirschbaum, M.U.F., 1994. The sensitivity of C[3] photosynthesis to increasing CO[2] concentration : a theoretical analysis of its dependence on temperature and background CO[2] concentration. *Plant, Cell and Environment* 17, 747-754.

Kirschbaum, M.U.F., Küppers, M., Schneider, H., Giersch, C., Noe, S., 1997. Modelling photosynthesis in fluctuating light with inclusion of stomatal conductance, biochemical activation and pools of key photosynthetic intermediates. *Planta* 204, 16-26.

Klamt, S., Stelling, J., 2003. Two approaches for metabolic pathway analysis? *Trends in Biotechnology* 21, 64-69.

Klamt, S., Stelling, J., Ginkel, M., Gilles, E.D., 2003. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* 19, 261-269.

Krab, K., 1995. Kinetic and regulatory aspects of the function of the alternative oxidase in plant respiration. *Journal of Bioengineering and Biomembranes* 27, 387-396.

Kruckeberg, A.L., Neuhaus, H.E., Feil, R., Gottlieb, L.D., Stitt, M., 1989. Decreased-activity mutants of phosphoglucose isomerase in the cytosol and chloroplast of *Clarkia xantiana*. Impact on mass-action ratios and fluxes to sucrose and starch, and estimation of Flux Control Coefficients and Elasticity Coefficients. *Biochemical Journal* 261, 457-467.

Laisk, A., Edwards, G.E., 2000. A mathematical model of C(4) photosynthesis: the mechanism of concentrating CO(2) in NADP-malic enzyme type species. *Photosynthesis Research* 66, 199-224.

Lange, B.M., 2006. Integrative analysis of metabolic networks: from peaks to flux models? *Current Opinion in Plant Biology* 9, 220-226.

Lange, B.M., Ghassemian, M., 2006. Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry* 66, 413-451.

Latowski, D., Burda, K., Strzalka, K., 2000. A mathematical model describing kinetics of conversion of violaxanthin to zeaxanthin via intermediate antheraxanthin by the xanthophyll cycle enzyme violaxanthin de-epoxidase. *Journal of Theoretical Biology* 206, 507-514.

Leduc, M., Tikhomiroff, C., Cloutier, M., Perrier, M., Jolicoeur, M., 2006. Development of a kinetic metabolic model: application to *Catharanthus roseus* hairy root. *Bioprocess and Biosystems Engineering* 28, 295-313.

Lee, D.Y., Yun, C., Cho, A., Hou, B.K., Park, S., Lee, S.Y., 2006. WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. *Bioinformatics* 22, 1150-1151.

Le Novere, N., Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J.L., Spence, H.D., Wanner, B.L., 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology* 23, 1509-1515.

- Le Novere, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L., Hucka, M., 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research* 34, D689-D691.
- Malone, J.G., Mittova, V., Ratcliffe, R.G., Kruger, N.J., 2006. The response of carbohydrate metabolism in potato tubers to low temperature. *Plant and Cell Physiology* 47, 1309-1322.
- Martinez-Force, E., Garces, R., 2000. Metabolic control analysis of *de novo* sunflower fatty acid biosynthesis. *Biochemical Society Transactions* 28, 669-671.
- Matsuda, F., Morino, K., Miyashita, M., Miyagawa, H., 2003. Metabolic flux analysis of the phenylpropanoid pathway in wound-healing potato tuber tissue using stable isotope-labeled tracer and LC-MS spectroscopy. *Plant and Cell Physiology* 44, 510-517.
- Matsuda, F., Morino, K., Ano, R., Kuzawa, M., Wakasa, K., Miyagawa, H., 2005. Metabolic flux analysis of the phenylpropanoid pathway in elicitor-treated potato tuber tissue. *Plant and Cell Physiology* 46, 454-466.
- McNeil, S.D., Nuccio, M.L., Rhodes, D., Shachar-Hill, Y., Hanson, A.D., 2000a. Radiotracer and computer modeling evidence that phospho-base methylation is the main route of choline synthesis in tobacco. *Plant Physiology* 123, 371-380.
- McNeil, S.D., Rhodes, D., Russell, B.L., Nuccio, M.L., Shachar-Hill, Y., Hanson, A.D., 2000b. Metabolic modeling identifies key constraints on an engineered glycine betaine synthesis pathway in tobacco. *Plant Physiology* 124, 153-162.
- Mendes, P., 1993. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Computer Applications in the Biosciences* 9, 563-571.
- Mendoza-Cózatl, D.G., and Moreno-Sánchez, R., 2006. Control of glutathione and phytochelatin synthesis under cadmium stress. *Pathway modeling for plants. Journal of Theoretical Biology* 238, 919-936.
- Morgan, J.A., Rhodes, D., 2002. Mathematical modeling of plant metabolic pathways. *Metabolic Engineering* 4, 80-89.
- Neff, R., Blasius, B., Beck, F., Lüttge, U., 1998. Thermodynamics and energetics of the tonoplast membrane operating as a hysteresis switch in an oscillatory model of Crassulacean acid metabolism. *Journal of Membrane Biology* 165, 37-43.
- Nuccio, M.L., McNeil, S.D., Ziemak, M.J., Hanson, A.D., Jain, R.K., Selvaraj, G., 2000. Choline import into chloroplasts limits glycine betaine synthesis in tobacco: analysis of

plants engineered with a chloroplastic or a cytosolic pathway. *Metabolic Engineering* 2, 300-311.

Nungesser, D., Kluge, M., Tolle, H., and Oppelt, W., 1984. A dynamic computer model of the metabolic and regulatory processes in Crassulacean acid metabolism. *Planta* 162, 204-214.

Olivier, B.G., Snoep, J.L., 2004. Web-based kinetic modelling using JWS Online. *Bioinformatics* 20, 2143-2144.

Padovan, A.C., Dry, I.B., Wiskich, J.T., 1989. An analysis of the control of phosphorylation-coupled respiration in isolated plant mitochondria. *Plant Physiology* 90, 928-933.

Page, R.A., Okada, S., Harwood, J.L., 1994. Acetyl-CoA carboxylase exerts strong flux control over lipid biosynthesis in plants. *Biochimica et Biophysica Acta* 1210, 369-372.

Papin, J.A., Price, N.D., Palsson, B.Ø., 2002. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Research* 12, 1889-1900.

Pearcy, R.W., Gross, L.J., He, D., 1997. An improved dynamic model of photosynthesis for estimation of carbon gain in sunfleck light regimes. *Plant, Cell and Environment* 20, 411-424.

Pettersson, G., 1997. Control properties of the Calvin photosynthesis cycle at physiological carbon dioxide concentrations *Biochimica et Biophysica Acta* 1322, 173-182

Pettersson, G., Ryde-Pettersson, U., 1988. A mathematical model of the Calvin photosynthesis cycle. *European Journal of Biochemistry* 175, 661-672.

Polle, A., 2001. Dissecting the superoxide dismutase-ascorbate-glutathione-pathway in chloroplasts by metabolic modeling. Computer simulations as a step towards flux analysis. *Plant Physiology* 126, 445-462.

Poolman, M.G., Fell, D.A., Thomas, S., 2000. Modelling photosynthesis and its control. *Journal of Experimental Botany* 51, 319-328.

Poolman, M.G., Fell, D.A., Raines, C.A., 2003. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *European Journal of Biochemistry* 270, 430-439.

Price, N.D., Reed, J.L., Papin, J.A., Famili, I., Palsson, B.Ø., 2003 Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophysical Journal* 84, 794-804.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews in Microbiology* 2, 886-897.

Quick, W.P., Schurr, U., Scheibe, R., Schulze, E.D., Rodermel, S.R., Bogorad, L., Stitt, M., 1991. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with "antisense" rbcS. 1. Impact on photosynthesis in ambient growth conditions. *Planta* 183, 542-554.

Ramli, U.S., Salas, J.J., Quant, P.A., Harwood, J.L., 2002. Control analysis of lipid biosynthesis in tissue cultures from oil crops shows that flux control is shared between fatty acid synthesis and lipid assembly. *Biochemical Journal* 364, 393-401.

Ramli, U.S., Salas, J.J., Quant, P.A., Harwood, J.L., 2005. Metabolic control analysis reveals an important role for diacylglycerol acyltransferase in olive but not in oil palm lipid accumulation. *FEBS Journal* 272, 5764-5770.

Ramsey, S., Orrell, D., Bolouri, H., 2005. Dizzy: stochastic simulation of large-scale genetic regulatory networks. *Journal of Bioinformatics and Computational Biology* 3, 415-436.

Ratcliffe, R.G., Shachar-Hill, Y., 2001. Probing plant metabolism with NMR. *Annual Review of Plant Physiology and Plant Molecular Biology* 52, 499-526.

Ratcliffe, R.G., Shachar-Hill, Y., 2006. Measuring multiple fluxes through plant metabolic networks. *The Plant Journal* 45, 490-511.

Reder, C., 1988. Metabolic control theory: a structural approach. *Journal of Theoretical Biology* 135, 175-201.

Roessner-Tunali, U., Liu, J., Leisse, A., Balbo, I., Perez-Melis, A., Willmitzer, L., Fernie, A.R., 2004. Kinetics of labelling of organic and amino acids in potato tubers by gas chromatography-mass spectrometry following incubation in (13)C labelled isotopes. *The Plant Journal* 39, 668-679.

Rohwer, J.M., Botha, F.C., 2001. Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. *Biochemical Journal* 358, 437-445.

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. *The Plant Journal* 42, 69-83.

Rontein, D., Dieuaide-Noubhani, M., Dufourc, E.J., Raymond, P., Rolin, D., 2002. The metabolic architecture of plant cells. Stability of central metabolism and flexibility of anabolic pathways during the growth cycle of tomato cells. *Journal of Biological Chemistry* 277, 43948-43960.



Roscher, A., Emsley, L., Raymond, P., Roby, C., 1998. Unidirectional steady state rates of central metabolism enzymes measured simultaneously in a living plant tissue. *Journal of Biological Chemistry* 273, 25053-25061.

Roscher, A., Kruger, N.J., Ratcliffe, R.G., 2000. Strategies for metabolic flux analysis in plants using isotope labelling. *Journal of Biotechnology* 77, 81-102.

Runquist, M., Kruger, N.J., 1999. Control of gluconeogenesis by isocitrate lyase in endosperm of germinating castor bean seedlings. *The Plant Journal* 19, 423-431.

SBML Forum, 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524-531.

Schilling, C.H., Edwards, J.S., Palsson, B.Ø., 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnology Progress* 15, 288-295.

Schilling, C.H., Letscher, D., Palsson, B.Ø., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology* 203, 229-248.

Schilling, C.H., Edwards, J.S., Letscher, D., Palsson, B.Ø., 2001. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering* 71, 286-306.

Schmidt, K., Carlsen, M., Nielsen, J., Villadsen, J., 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering* 55, 831-840.

Schmidt, K., Norregaard, L.C., Pedersen, B., Meissner, A., Duus, J.O., Nielsen, J.O., Villadsen, J., 1999. Quantification of intracellular metabolic fluxes from fractional enrichment and  $^{13}\text{C}$ - $^{13}\text{C}$  coupling constraints on the isotopomer distribution in labeled biomass components. *Metabolic Engineering* 1, 166-179.

Schmidt, H., Jirstrand, M., 2006. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* 22, 514-515.

Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems* 2, 165-182.

Schuster, S., Dandekar, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* 17, 53-60.

- Schwarz, R., Musch, P., von Kamp, A., Engels, B., Schirmer, H., Schuster, S., Dandekar, T., 2005. YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities. *BMC Bioinformatics* 6, 135.
- Schwender, J., Ohlrogge, J.B., 2002. Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiology* 130,347-361.
- Schwender, J., Ohlrogge, J.B., Shachar-Hill, Y., 2003. A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos. *Journal of Biological Chemistry* 278, 29442-29453.
- Schwender, J., Ohlrogge, J., Shachar-Hill, Y., 2004a. Understanding flux in plant metabolic networks. *Current Opinion in Plant Biology* 7, 309-317.
- Schwender, J., Goffman, F., Ohlrogge, J.B., Shachar-Hill, Y., 2004b. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432,779-782.
- Schwender, J., Shachar-Hill, Y., Ohlrogge, J.B., 2006. Mitochondrial metabolism in developing embryos of *Brassica napus*. *Journal of Biological Chemistry* 281, 34040-34047.
- Selivanov, V.A., Marin, S., Lee, P.W., Cascante, M., 2006. Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. *Bioinformatics* 22, 2806-2812.
- Sielewiesiuk, J., Gruszecki, W.I., 1991. A simple model describing the kinetics of the xanthophyll cycle. *Biophysical Chemistry* 41, 125-129.
- Slater, S., Mitsky, T.A., Houmiel, K.L., Hao, M., Reiser, S.E., Taylor, N.B., Tran, M., Valentin, H.E., Rodriguez, D.J., Stone, D.A., Padgett, S.R., Kishore, G., Gruys, K.J., 1999. Metabolic engineering of *Arabidopsis* and *Brassica* for poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. *Nature Biotechnology* 17, 1011-1016.
- Spielbauer, G., Margl, L., Hannah, L.C., Romisch, W., Ettenhuber, C., Bacher, A., Gierl, A., Eisenreich, W., Genschel, U., 2006. Robustness of central carbohydrate metabolism in developing maize kernels. *Phytochemistry* 67, 1460-1475.
- Sriram, G., Shanks, J.V., 2004. Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping. *Metabolic Engineering* 6, 116-132.
- Sriram, G., Fulton, D.B., Iyer, V.V., Peterson, J.M., Zhou, R., Westgate, M.E., Spalding, M.H., Shanks, J.V., 2004. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional (<sup>13</sup>C

labeling, two-dimensional [(13)C, (1)H] nuclear magnetic resonance, and comprehensive isotopomer balancing. *Plant Physiology* 136, 3043-3057.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D., 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190-193.

Stephanopoulos, G., Nielsen, J.B., Aristidou, A., 1998. *Metabolic Engineering: Principles and Methodologies*. Academic Press, San Diego.

Steuer, R., Gross, T., Selbig, J., Blasius, B., 2006. Structural kinetic modeling of metabolic networks. *Proceeding of the National Academy of Sciences of the United States of America* 103, 11868-11873.

Stitt, M., Quick, W.P., Schurr, U., Schulze, E.D., Rodermel, S.R., Bogorad, L., 1991. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with 'antisense' rbcS. 2. Flux control coefficients for photosynthesis in varying light, CO<sub>2</sub>, and air humidity. *Planta* 183, 555-566.

Sweetlove, L.J., Fernie, A.R., 2005. Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. *New Phytologist* 168, 9-24.

Szyperski, T., 1998. <sup>13</sup>C-NMR, MS and metabolic flux balancing in biotechnology research. *Quarterly Reviews in Biophysics* 31, 41-106.

Takahashi, K., Ishikawa, N., Sadamoto, Y., Sasamoto, H., Ohta, S., Shiozawa, A., Miyoshi, F., Naito, Y., Nakayama, Y., Tomita, M., 2003. E-CELL2: Multi-platform E-CELL simulation system. *Bioinformatics* 19, 1727-1729.

Teusink, B., Passarge, J., Reijenga, C.A., Esgalhad, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V., Snoep, J.L., 2000. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry* 267, 5313-5329.

Thomas, S., Mooney, P.J.F., Burrell, M.M., Fell, D.A., 1997a. Finite change analysis of glycolytic intermediates in tuber tissue of lines of transgenic potato (*Solanum tuberosum*) overexpressing phosphofructokinase. *Biochemical Journal* 322, 111-117.

Thomas, S., Mooney, P.J.F., Burrell, M.M., Fell, D.A., 1997b. Metabolic control analysis of glycolysis in tuber tissue of potato (*Solanum tuberosum*): explanation of the low control coefficient of phosphofructokinase over respiratory flux. *Biochemical Journal* 322, 119-127.

Urbanczik, R., 2006. SNA - a toolbox for the stoichiometric analysis of metabolic networks. *BMC Bioinformatics* 13, 129.

van Winden, W., Heijnen, J.J., Verheijen, P.J.T., 2002. Cumulative bondomers: a new concept in flux analysis from  $^2\text{D}$ ,  $^{13}\text{C}$ ,  $^1\text{H}$ -COSY NMR data. *Biotechnology and Bioengineering* 80, 731-745.

Varner, J., Ramkrishna, D., 1999. Mathematical models of metabolic pathways. *Current Opinion in Biotechnology* 10, 146-150.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* 22, 1930-1931.

Wang, Z., Zhu, X.G., Chen, Y., Li, Y., Hou, J., Li, Y., Liu, L., 2006. Exploring photosynthesis evolution by comparative analysis of metabolic networks between chloroplasts and photosynthetic bacteria. *BMC Genomics* 7, 100.

Whittaker, A., Botha, F.C., 1997. Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. *Plant Physiology* 115, 1651-1659.

Wiechert, W., 2001.  $^{13}\text{C}$  metabolic flux analysis. *Metabolic Engineering* 3, 195-206.

Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., de Graaf, A.A., 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and Bioengineering* 66, 69-85.

Wiechert, W., Möllney, M., Petersen, S., de Graaf, A.A., 2001. A universal framework for  $^{13}\text{C}$  metabolic flux analysis. *Metabolic Engineering* 3, 265-283.

Wiechert, W., Nöh, K., 2005. From stationary to instationary metabolic flux analysis. *Advances in Biochemical Engineering and Biotechnology* 92, 145-172.

Wiechert, W., 2002. Modeling and simulation: tools for metabolic engineering. *Journal of Biotechnology* 94, 37-63.

Xia, Y., Yu, H., Jansen, R., Seringhaus, M., Baxter, S., Greenbaum, D., Zhao, H., Gerstein, M., 2004. Analyzing cellular biochemistry in terms of molecular networks. *Annual Review of Biochemistry* 73, 1051-1087.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P., Potrykus, I., 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287, 303-305.

You, L., Hoonlor, A., Yin, J., 2003. Modeling biological systems using Dynetica--a simulator of dynamic networks. *Bioinformatics* 19, 435-436.

Yun, H., Lee, D.Y., Jeong, J., Lee, S., Lee, S.Y., 2005. MFAML: a standard data structure for representing and exchanging metabolic flux models. *Bioinformatics* 21, 3329-3330.

Zamboni, N., Fischer, E., Sauer, U., 2005. FiatFlux - a software for metabolic flux analysis from  $^{13}\text{C}$ -glucose experiments. *BMC Bioinformatics* 25, 209.

Zimmer, W., Brüggemann, N., Emeis, S., Giersch, C., Lehning, A., Steinbrecher, R., Schnitzler, J.P., 2000. Process-based modelling of isoprene emission by oak leaves. *Plant, Cell and Environment*, 23 585-595.

Zupke, C., Stephanopoulos, G., 1994. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnology Progress* 10, 489-498.

## CHAPTER 3

### **A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint**

Rigoberto Rios-Esteva<sup>†,¶</sup>, Glenn W. Turner<sup>†</sup>, James M. Lee<sup>¶</sup>, Rodney B. Croteau<sup>†,\*</sup>, and B. Markus Lange<sup>†,‡,§,\*</sup>

<sup>†</sup>Institute of Biological Chemistry, <sup>‡</sup>M.J. Murdock Metabolomics Laboratory, <sup>§</sup>Center for Integrated Biotechnology, and <sup>¶</sup>School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA 99164-6340

**Author contributions:** RRE, GWT and BML designed research; RRE and GWT performed research; RRE, GWT, JML, RBC and BML analyzed data; and RRE, RBC and BML wrote the paper.

**Abbreviations:** MFS, (+)-menthofuran synthase; PR, (+)-pulegone reductase.

\* To whom correspondence should be addressed. E-mails: croteau@wsu.edu, lange-m@wsu.edu.

Reproduced with permission from “*Proceedings of the National Academy of Sciences of the USA* 2008 February 26; 105(8): 2818–2823.

Copyright © 2008 By The National Academy of Sciences of the USA

### **Abstract**

The integration of mathematical modeling and experimental testing is emerging as a powerful approach for improving our understanding of the regulation of metabolic pathways. In this study, we report on the development of a kinetic mathematical model that accurately simulates the developmental patterns of monoterpenoid essential oil accumulation in peppermint (*Mentha x piperita*). This model was then utilized to evaluate the biochemical processes underlying experimentally determined changes in the monoterpene pathway under low ambient light intensities, which lead to an accumulation of the branchpoint intermediate (+)-pulegone and the side product (+)-menthofuran. Our simulations indicated that the environmentally-regulated changes in monoterpene profiles

could only be explained when, in addition to effects on biosynthetic enzyme activities, as yet unidentified inhibitory effects of (+)-menthofuran on the branchpoint enzyme pulegone reductase (PR) were assumed. Subsequent *in vitro* analyses with recombinant protein confirmed that (+)-menthofuran acts as a weak competitive inhibitor of PR ( $K_i = 300 \mu\text{M}$ ). To evaluate if intracellular concentration of (+)-menthofuran were high enough for PR inhibition *in vivo*, we isolated essential oil-synthesizing secretory cells from peppermint leaves and subjected them to steam distillations. When peppermint plants were grown under low light conditions, (+)-menthofuran was selectively retained in secretory cells and accumulated to very high levels (up to 20 mM), whereas under regular growth conditions (+)-menthofuran levels remained very low (below 400  $\mu\text{M}$ ). These results illustrate the utility of iterative cycles of mathematical modeling and experimental testing to elucidate the mechanisms controlling flux through metabolic pathways.

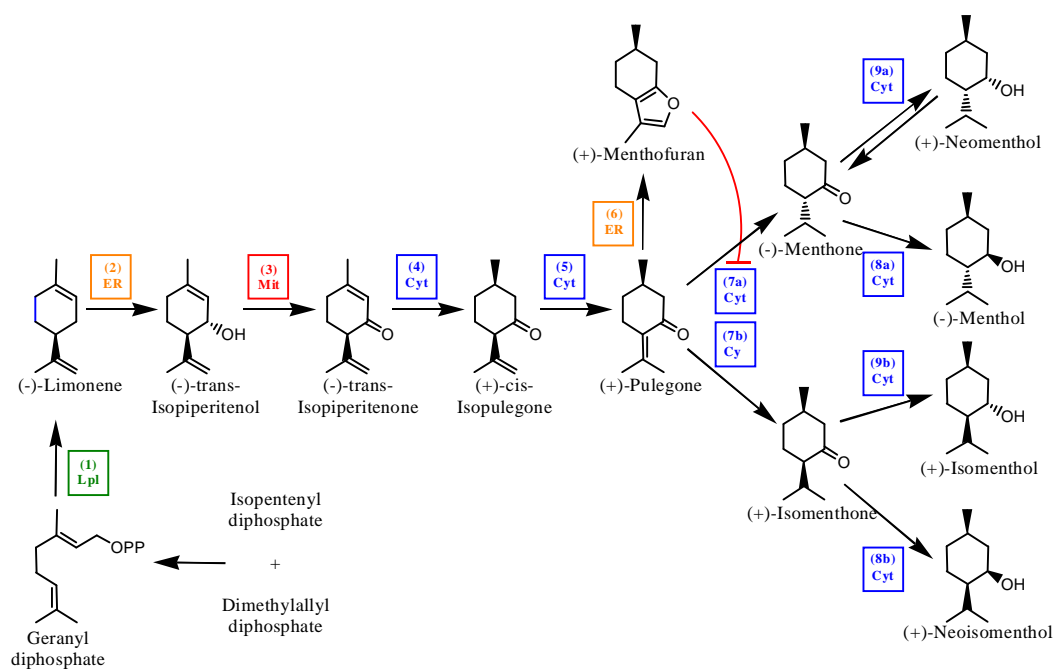
### 3.1. Introduction

The commercially valuable essential oil of peppermint (*Mentha x piperita*) consists primarily of *p*-menthane-type monoterpenes [1], which are synthesized and accumulated in leaf protuberances termed peltate glandular trichomes [2, 3]. Within these trichomes, the biosynthesis of monoterpenes is restricted to non-photosynthetic secretory cells that are arranged in an eight-celled disk [4]. The secretory cells exude essential oil into an emerging cavity formed by the separation of a preformed layer of cuticular material [5]. Modification of a general protocol for the isolation of peppermint

secretory cells [6] enabled the extraction of high quality mRNA, the generation of cDNA libraries, the sequencing of randomly selected clones, and the functional testing of cDNAs by expression of recombinant proteins in microbial hosts [7]. In combination with developmental studies at the microscopic, molecular, and biochemical levels, these functional genomics approaches have yielded a wealth of information about the biochemical properties of the individual biosynthetic enzymes and the regulation of the monoterpene pathway as a whole (reviewed in [8]).

In peppermint secretory cells, the precursors of monoterpenes are derived exclusively from the leucoplast-localized mevalonate-independent pathway [9, 10]. The first committed step of the monoterpene pathway, the conversion of geranyl diphosphate to (-)-limonene, is catalyzed by (-)-limonene synthase [11, 12], which is also localized to leucoplasts [13] (Fig. 3.1). Following translocation to the endoplasmic reticulum, (-)-limonene undergoes a cytochrome P450-dependent hydroxylation, catalyzed by (-)-limonene 3-hydroxylase [14], to form (-)-trans-isopiperitenol. After import into mitochondria, (-)-trans-isopiperitenol is oxidized to (-)-isopiperitenone by a specific NAD<sup>+</sup>-dependent short-chain dehydrogenase [15, 16]. A double-bond reduction, catalyzed by (-)-isopiperitenone reductase [17], followed by an isomerization of (+)-cis-isopulegone [18] generate the branchpoint intermediate (+)-pulegone in the cytosol. Cytosolic (+)-pulegone reductase (PR) synthesizes both (-)-menthone and (+)-isomenthone from (+)-pulegone [17]. Further conversions in the main monoterpene pathway are catalyzed by the bifunctional cytosolic activities of (-)-menthone:(-)-(3*R*)-menthol reductase and (-)-menthone:(-)-(3*S*)-neomenthol reductase [19].





**Fig. 3.1.** Outline of *p*-menthane monoterpene metabolism in peppermint glandular trichomes. The following enzymes are involved in this pathway: (1), (-)-limonene synthase; (2), (-)-limonene 3-hydroxylase; (3), (-)-trans-isopiperitenol dehydrogenase; (4), (-)-trans-isopiperitenone reductase; (5), (+)-cis-isopulegone isomerase; (6), (+)-menthofuran synthase; (7a), (+)-pulegone reductase ((-)-menthone-forming activity); (7b), (+)-pulegone reductase ((+)-isomenthone-forming activity); (8a), (-)-menthone: (-)-menthol reductase ((-)-menthol-forming activity); (8b), (-)-menthone: (-)-menthol reductase ((+)-neoisomenthol-forming activity); (9a), (-)-menthone: (+)-neomenthol reductase ((+)-neomenthol-forming activity); (9b), (-)-menthone: (+)-neomenthol reductase ((+)-isomenthol-forming activity). The subcellular compartmentation of *p*-menthane metabolic enzymes is color-coded: Cyt (blue), cytosol; ER (orange), endoplasmic reticulum; Lpl (green), leucoplasts; Mit (red), mitochondria. The inhibition of (+)-pulegone reductase by (+)-menthofuran, as demonstrated in the present manuscript, is indicated by a red arc with an orthogonal red line.

The side product (+)-menthofuran can be formed from (+)-pulegone by the action of (+)-menthofuran synthase (MFS), an endoplasmic reticulum-localized cytochrome P450-dependent monooxygenase [20]. The highest rates of monoterpene biosynthesis were measured during the period of maximal leaf expansion (12 to 20 d after leaf initiation) [21]. Based on  $^{14}\text{CO}_2$  incorporation experiments with subsequent radio-GC analysis, the rate of monoterpene biosynthesis appeared to be the most important factor controlling leaf monoterpene levels. Subsequent studies established that the transcript levels of genes involved in the peppermint monoterpene pathway and the corresponding enzyme activities correlated well with the rate of monoterpene biosynthesis [22], indicating that flux might be coordinately regulated at the level of gene expression.

High quality oils are characterized by a complex compositional balance of monoterpenes with high (-)-menthol, moderate (-)-menthone, and low (+)-pulegone and (+)-menthofuran quantities [23, 24]. Under adverse environmental conditions, such as low light intensity, water deficit and/or high night temperatures, elevated quantities of (+)-pulegone and (+)-menthofuran accumulate [25, 26], thus rendering an oil of unsatisfactory quality. Mahmoud and Croteau reported that transgenic plants with decreased (+)-menthofuran synthase transcript levels accumulated vastly reduced (+)-pulegone and (+)-menthofuran amounts under regular and stress conditions [27]. Further studies indicated that PR transcript levels decreased in the presence of (+)-menthofuran, thus resulting in a decreased PR activity and increased (+)-pulegone amounts [28]. Because of these regulatory complexities and the occurrence of branchpoints, the fine-tuning of monoterpene biosynthesis can not be understood intuitively.

Here we report on the development of a kinetic mathematical model, based upon the available experimental data, that accurately describes the behavior of the peppermint monoterpene biosynthetic pathway under various experimental conditions. Model predictions were used to generate non-trivial, testable hypotheses regarding poorly understood regulatory mechanisms, and modeling-guided follow-up experiments were used to demonstrate an as yet unidentified role for (+)-menthofuran as a competitive inhibitor of PR. These results indicate that gene expression and posttranslational modulation of enzyme activity are both important factors in regulating peppermint monoterpene biosynthesis.

## **3.2 Results and Discussion**

### **3.2.1. Development of a mathematical model simulating monoterpene biosynthesis in peppermint oil gland secretory cells.**

The filling of peppermint glandular trichomes with monoterpenoid essential oil is a complex process. A mathematical model simulating this process needs to account for various levels of regulation. Several estimates were made to allow linking microscopic variables (e.g., monoterpene composition in individual glandular trichomes) with macroscopic measurements (e.g., leaf monoterpene composition). The number of glandular trichomes per leaf depends on environmental and developmental parameters. Under greenhouse conditions the number of biosynthetically active glandular trichomes increases from about 2,500 (at day 5 after leaf initiation) to about 13,000 (at day 18 after initiation) ([29] and our own unpublished data), whereas under reduced light intensity

( $300 \mu\text{mol m}^{-2} \text{ s}^{-1}$ ) the maximum number of glandular trichomes is about 7,500 [30] (details in Supporting Information). The volume of the cluster of eight secretory cells of each individual glandular trichome, which is not dependent on environmental conditions, can be determined based on microscopic size measurements (average diameter of the secretory cell disk is  $60 \mu\text{M}$ , height is  $16 \mu\text{M}$ ). We approximated the shape of the secretory cell cluster as a frustum of a cone, the volume of which can be calculated as  $\frac{1}{3} \pi h (R^2 + R r + r^2)$ , thus resulting in a calculated average secretory cell disk volume of  $2.35 \cdot 10^{-5} \mu\text{L}$  at maturity. Morphometric measurements using microscopic images at different developmental stages were combined with stereological approaches [31] to calculate the volume densities of subcellular compartments in peppermint oil gland secretory cells, thus allowing us to estimate concentrations of enzymes based upon prior knowledge regarding their organellar distribution (details in Supporting Information). We used two independent methods to calculate the amounts of monoterpenes produced per individual glandular trichome. First, monoterpene amounts obtained from GC-FID analyses of steam-distilled leaves were divided by the number of oil-bearing glandular trichomes. Second, the volume of the essential oil-filled subcuticular cavity of mature glandular trichomes was calculated based on the approximation of its shape as a hemisphere ( $\frac{2}{3} \pi r^3$ ) with a diameter of  $65 \mu\text{M}$  (determined with morphometric measurements using microscopic images), from which the volume of the secretory cell cluster was subtracted. The volume ( $7.53 \cdot 10^{-5} \mu\text{L}$ ) was then multiplied by the known essential oil density ( $0.9 \text{ mg} / \mu\text{L}$ ) to obtain the monoterpene amount per gland. An average monoterpene molecular weight of  $150 \text{ g} / \text{mol}$  was assumed for converting these values into molar amounts per gland (for details see Supporting Information). Both

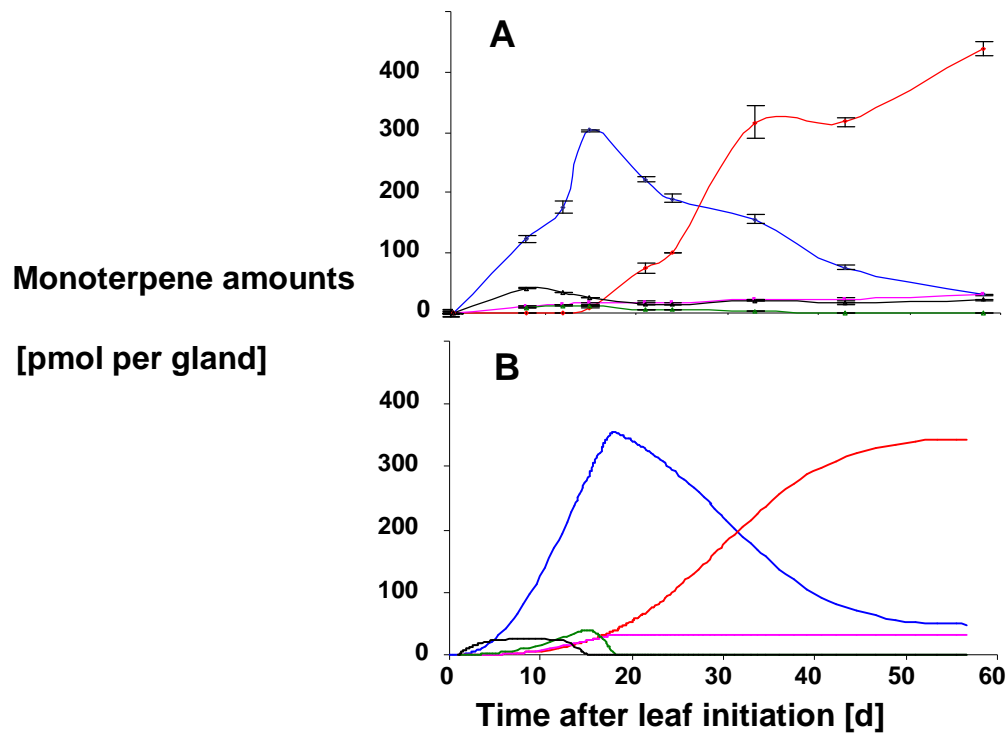
approaches yielded very similar results with monoterpene amounts between 470 and 570 pmol per gland at maturity, indicating that the essential oil in peppermint glandular trichomes consists almost exclusively of monoterpenes.

Kinetic constants for monoterpene biosynthetic enzymes were obtained from the literature [11, 15, 17-19, 32]. Michaelis-Menten rate equations were used to describe the kinetic behavior of individual biosynthetic enzymes. These expressions were used in a set of ordinary differential equations to account for the time dependence of the metabolite concentration (details in Supporting Information). In addition, we accounted for changes in enzyme concentration during the course of leaf development by approximating the shape of previously reported enzyme activity patterns [22] with Gauss functions (details in Supporting Information). We did not account for diurnal changes in enzyme activities and the effects of day/night temperature changes on enzyme kinetics; thus, the kinetic and enzyme activity values in our model represent ‘daily averages’. A dynamic simulation of monoterpene profiles was performed by simultaneously solving the system of modified ordinary differential equations (expressing both enzyme properties and expression patterns) using the ‘ode15’ function of the MATLAB software package. Since enzyme concentrations and the variables used for the Gauss function representing developmental enzyme activity patterns could only be estimated, iterative optimizations of the initial model parameters were performed, using the experimentally measured monoterpene profiles as constraints, until the best fit of modeling results and experimental data was achieved. Our modeling simulated an early accumulation of high levels of (-)-menthone (maximum at 15 d), which is converted to (-)-menthol during the essential oil maturation phase (15 - 55 d), a transient low-level accumulation of (-)-limonene and (+)-pulegone,

and a hyperbolic time course of (+)-menthofuran accumulation (low levels). These simulations were in excellent agreement with monoterpene profiles obtained with greenhouse-grown plants (Fig. 3.2). The fact that monoterpene profiles under regular growth conditions could be simulated successfully by considering only the kinetic properties and developmental expression patterns of biosynthetic enzymes confirmed previous conclusions that monoterpene biosynthesis is determined by the rate of biosynthesis [21], most likely regulated at the level of gene expression [22].

### **3.2.2. Mathematical modeling suggests that post-translational regulation determines monoterpene profiles under environmental stress conditions.**

Peppermint oil of the highest commercial quality can only be produced with plants grown in certain geographical regions with hot days and cool nights. It has also been known for many years that essential oil yield and composition vary widely among growing regions and are affected by numerous environmental and agronomic factors (reviewed in [33]). To evaluate the suitability of mathematical modeling for understanding environmental effects on essential oil biosynthesis, we subjected peppermint plants to a series of environmental stresses, measured monoterpene profiles, and tested several hypotheses regarding biochemical and regulatory mechanisms underlying the accumulation of undesirable essential oil components, in particular (+)-pulegone and (+)-menthofuran, based on the degree of convergence between these experimental data and computational simulations.

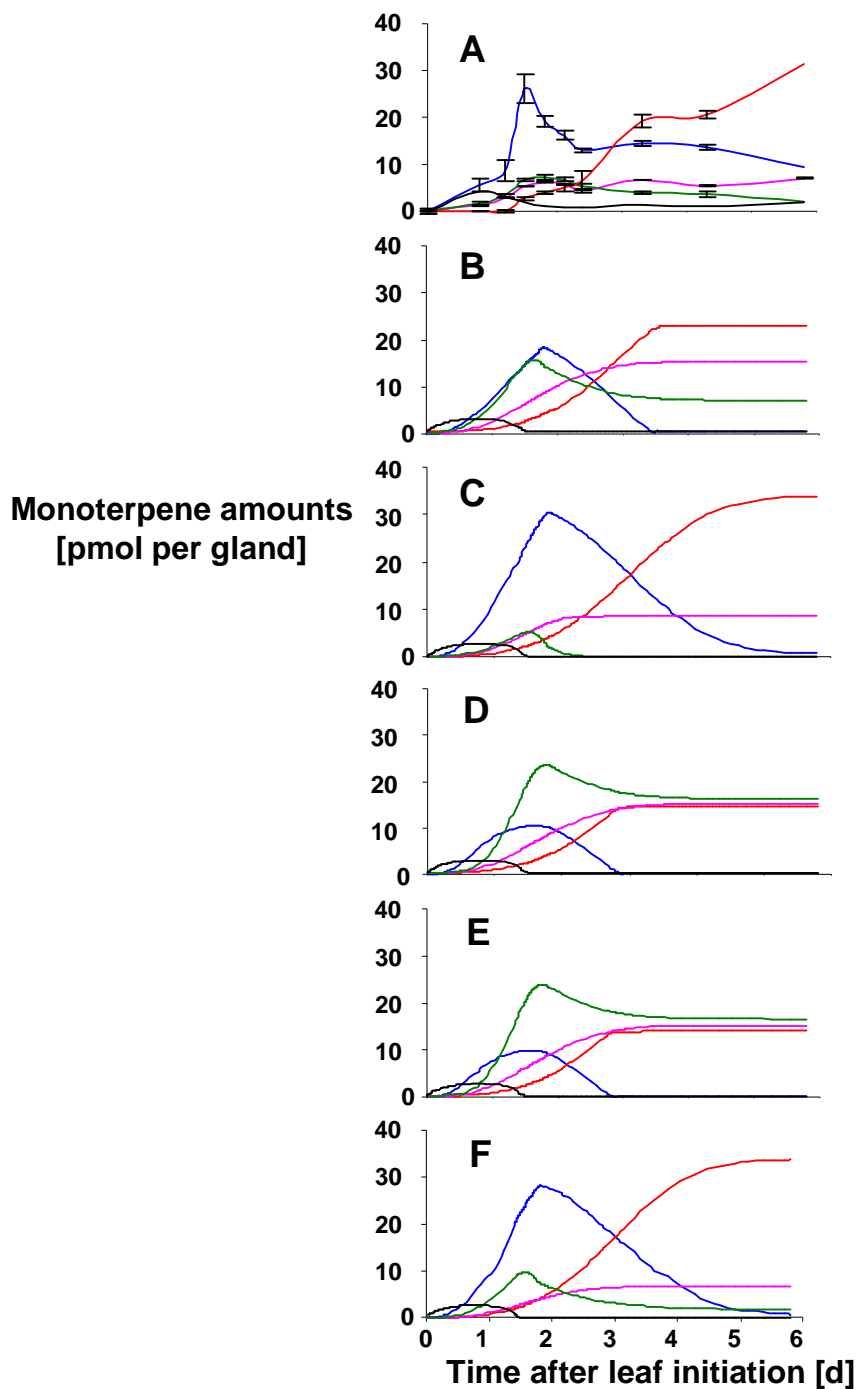


**Fig. 3.2.** Experimentally determined monoterpene profiles of peppermint plants grown under greenhouse conditions (**A**) and computer simulation based upon a kinetic mathematical model of mint monoterpene biosynthesis (**B**). The following colors are used for indicating monoterpene profiles: (-)-limonene, black; (+)-pulegone, pink; (+)-menthofuran, green; (-)-menthone, blue; (-)-menthol, red.

Under all environmental conditions tested (reduction of water/fertilizer to 50 %, lowering of light intensity, or increase of night temperatures) a reduction of total oil yield was measured, which correlated with smaller leaves and a lower number of glandular trichomes per leaf (unpublished results). In certain experiments (low light and high night temperatures) we also detected an accumulation of (+)-pulegone and (+)-menthofuran (data not shown), the simulation of which required testing various hypotheses computationally. Here we are going to use the low light experiment as an example to illustrate the process of computational hypothesis testing.

Control plants were grown in a greenhouse with additional lighting from sodium vapor lights ( $850 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation), a 16 h photoperiod and a temperature cycle of 27°C/21°C (day/night), whereas the experimental treatment involved plants kept in a growth chamber at reduced light intensity ( $300 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation) but under otherwise identical conditions as controls. Total essential oil yield was about 50 % lower in plants grown under low light conditions (480-630  $\mu\text{g}$  per leaf at maturity) compared to controls (1,160 to 1,270  $\mu\text{g}$  per leaf at maturity) (Fig. 3.3A). Cultivation under low light conditions lead to a transient increase of (+)-pulegone, with a maximum of 70 pmol per glandular trichome at 18 d, and a hyperbolic accumulation of (+)-menthofuran to roughly 60 pmol per glandular trichome. Mahmoud and Croteau had previously reported that (+)-pulegone and (+)-menthofuran increased and decreased in concert under stress conditions [27, 28]. Stem feeding experiments with (+)-menthofuran lead to a dose-dependent decrease in the expression of the gene encoding (+)-pulegone reductase (PR), the enzyme responsible for the conversion of (+)-pulegone into (-)-menthone, by an as yet unidentified mechanism [28].





**Fig. 3.3.** Monoterpene profiles of peppermint plants maintained in growth chambers under low light conditions (**A**) and computer simulation considering a reduction of PR and increase in MFS transcript levels (as suggested in [28]) (**B**). Simulations assuming an inhibition of PR by (+)-menthofuran by a competitive (**C**), uncompetitive (**D**) or noncompetitive (**E**) mechanism. Simulation of monoterpene profiles under low light conditions after model optimization (**F**). Color-code of monoterpene profiles as in Fig. 3.2.

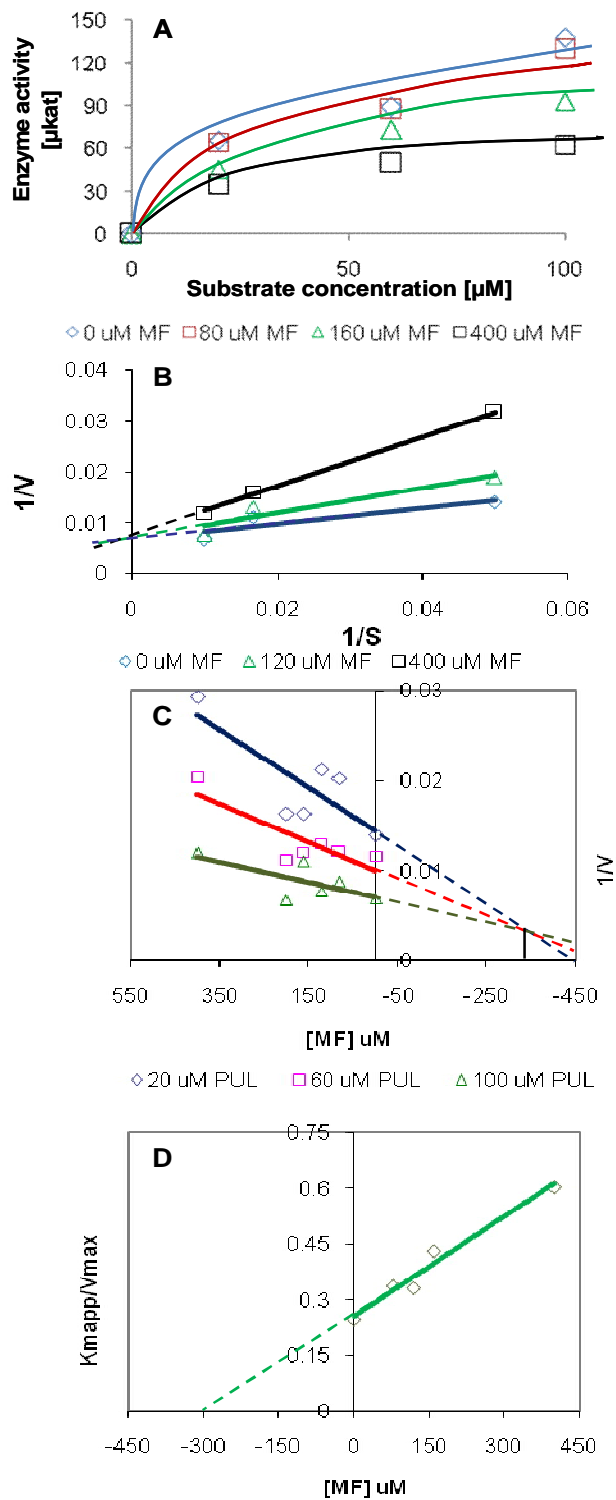
Furthermore, in transgenic lines with increased expression levels of the gene encoding (+)-menthofuran synthase (MFS) and higher (+)-menthofuran amounts in the essential oil, (+)-pulegone amounts were higher than in controls, which led to the hypothesis that the metabolic fate of (+)-pulegone is controlled via a (+)-menthofuran-mediated transcriptional down-regulation of PR levels [28]. To test if these assumptions could guide simulations of the monoterpene profiles observed in the present experiments with low light-grown plants, we initially increased the levels of MFS and decreased PR levels (2-fold up and 2-fold down, respectively) in our mathematical model (Supporting Information). Simulations using these model adjustments indicated that, compared to controls, the peak levels of (+)-pulegone and (+)-menthofuran should be dramatically increased (160 and 150 pmol per gland, respectively), whereas (-)-menthone and (-)-menthol levels would be drastically reduced (190 and 220 pmol per gland, respectively) (Fig. 3.3B). We then simulated numerous other reasonable combinations with increased MFS and decreased PR levels (1.5 to 2.5-fold up and down, respectively), but a satisfactory simulation of the measured monoterpene profiles could not be obtained. Since these simulations were only in partial agreement with experimentally determined values, alternative hypotheses regarding the biochemical mechanisms of environmental variation in peppermint essential oil composition had to be considered.

As a plausible and as yet unexplored possibility for peppermint monoterpene pathway regulation we simulated the essential oil composition if inhibitory effects of (+)-menthofuran on PR were assumed. Using our model we tested for competitive, uncompetitive and noncompetitive inhibition using assumed  $K_i$  values of 10  $\mu\text{M}$  (Matlab code in Supporting Information). Simulations assuming a competitive inhibition

mechanism were in excellent agreement with monoterpene values measured experimentally (Fig. 3.3.C), whereas simulations for uncompetitive or noncompetitive inhibition of PR by (+)-menthofuran predicted very high accumulation levels for (+)-pulegone and (+)-menthofuran (Fig. 3.3 D, E). Based on these computational predictions, we then tested experimentally if (+)-menthofuran exerted inhibitory effects on PR activity.

### **3.2.3 (+)-Menthofuran is a competitive inhibitor of (+)-pulegone reductase.**

Recombinant PR was expressed in *E. coli*, partially purified and assayed as described previously [17]. The kinetic constants we determined for (+)-pulegone as a substrate ( $K_m = 40 \mu\text{M}$ ,  $V_{max} = 185 \text{ pmol/s}$ , and  $\text{IC}_{50}$  (substrate inhibition) =  $150 \mu\text{M}$ ) were very similar to those reported previously, although our  $K_m$  value was a bit higher [17]. After completing these preliminary studies to establish the appropriate assay conditions, PR enzyme activity was measured with (+)-pulegone as a substrate (0 - 100  $\mu\text{M}$ ), NADPH as a cofactor (500  $\mu\text{M}$ ), and varying concentrations of the putative inhibitor (+)-menthofuran (0 - 400  $\mu\text{M}$ ). We observed a dose-dependent decrease in PR activity in the presence of (+)-menthofuran (Fig. 3.4A). To evaluate the mechanism of inhibition we used the Lineweaver-Burk method [34], in which  $1/V$  ( $V = \text{velocity}$ ) is plotted against  $1/[S]$  ( $[S] = \text{substrate concentration}$ ) (Fig. 3.4B). The lines obtained for different inhibitor concentrations had a common intercept with the  $1/V$  axis but the slopes of the lines increased with rising inhibitor concentrations. The  $V_{max}$  value (determined based on intercept with the  $1/V$  axis) remained the same in the presence of different



**Fig 3.4.** Characterization of (+)-menthofuran as a competitive inhibitor of peppermint PR. **(A)** Effect of various (+)-menthofuran concentration on PR activity. **(B)** Lineweaver-Burk plot for determining the type of inhibition exerted. Determination of the inhibition constant for (+)-menthofuran using a Dixon plot **(C)**.

inhibitor amounts, whereas the  $K_m$  value (determined based on the intercept with the  $1/[S]$  axis) increased with rising inhibitor concentrations, thus indicating a competitive inhibition mechanism. The inhibition constant ( $K_i$ ), which defines the competition of substrate and inhibitor for the same active site of PR, was determined using two independent approaches: with the Dixon method (plotting  $1/V$  against  $[I]$  ( $[I]$  = inhibitor concentration),  $K_i$  was determined by linear regression analysis [35]; with the  $K_{m(app)}$  method (plotting  $K_{m(app)}/V_{max}$  against  $[I]$ ),  $K_i$  was obtained by a nonlinear regression analysis [36]. In both cases a  $K_i$  for (+)-menthofuran as a competitive inhibitor of roughly 300  $\mu\text{M}$  was calculated. Using the same methods we estimated a  $K_i$  of 112  $\mu\text{M}$  for substrate inhibition by (+)-pulegone. Since (+)-menthofuran appeared to be a relatively weak competitive inhibitor ( $K_i$  value for (+)-menthofuran roughly 7.5-fold higher than the  $K_m$  value for (+)-pulegone as substrate), we tested if (+)-menthofuran concentrations in peppermint secretory cells were sufficiently high to give rise to relevant inhibitory effects.

#### **3.2.4 (+)-Menthofuran is preferentially retained in peppermint oil gland secretory cells.**

Leaves from greenhouse-grown peppermint plants were harvested at 20 d and 50 d, secretory cells were isolated and steam-distilled, and monoterpene profiles were analyzed by GC-FID (modified from [17]). Secretory cells from plants grown in the greenhouse contained primarily (-)-menthone at 20 d and substantial amounts of (-)-menthol at 50 d, whereas only small amounts of (+)-menthofuran and negligible amounts of (+)-pulegone were detected (details in Supporting Information). In contrast, secretory

cells obtained from plants grown under stress conditions (low light intensity) accumulated (+)-menthofuran as the principal metabolite (up to 20 mM, accounting to roughly 90 % of total monoterpenes). High amounts of (+)-menthofuran had been detected in secretory cells previously when grown under comparable growth conditions [3], but it had not been recognized that this was due to stress conditions (low light intensity). When we obtained essential oil directly from the subcuticular storage cavity of intact glandular trichomes using microcapillaries and analyzed it by GC-FID, the monoterpene profile was similar to that of steam-distilled whole leaves (low amounts of (+)-menthofuran in greenhouse-grown plants and moderate amounts in stressed plants) (data not shown). These results indicated that (+)-menthofuran was preferentially retained in secretory cells under stress conditions. Based on these new findings our mathematical model was updated to reflect the experimentally determined kinetic properties of PR ( $K_{m(+)-pulegone} = 40 \text{ uM}$ ;  $K_{i(+)-pulegone} = 112 \text{ uM}$  (by introducing a factor  $q$  the percentage of the the total PR activity affected by substrate inhibition can be adjusted);  $K_{i(+)-menthofuran} = 300 \text{ uM}$  (by introducing a factor  $p$  the percentage of the total PR activity affected by (+)-menthofuran inhibition can be adjusted)) and the high intracellular concentration of (+)-menthofuran (by introducing a factor  $z$  the local concentration of (+)-menthofuran can be adjusted). Simulations of low light conditions (Fig. 3.5E) were in excellent agreement with experimentally determined monoterpene profiles. Using this second generation model, different environmental conditions can now be simulated by simply modifying a set of variable factors. Our simulations also have the potential of enabling knowledge-based metabolic engineering approaches aimed at modulating monoterpene yield and composition. An improved understanding of the

transport processes involved in monoterpenoid essential oil biosynthesis and secretion will be a key future challenge for advancing our simulation efforts.

### **3.3 Materials and Methods**

#### **3.3.1 Plant Material.**

Peppermint (*Mentha x piperita* cv. Black Mitchum) plants were grown on soil (Sunshine Mix LC1, SunGro Horticulture) in a greenhouse with supplemental lighting from sodium vapor lights ( $850 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation at plant canopy level) with a 16 h photoperiod and a temperature cycle of 27°C/21°C (day/night). Plants were watered daily with a fertilizer mix (N:P:K 20:20:20, v/v/v; plus iron chelate and micronutrients). Monoterpene analyses were performed with leaves that were initiated on three-week old stems and were harvested at ages ranging from 5 to 55 days after bud formation. Stress experiments were performed by moving plants to a growth chamber with a 16 h photoperiod at reduced light levels ( $300 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation at plant canopy level).

#### **3.3.2 Monoterpene analysis.**

Leaves and isolated secretory cells [37] were directly (without prior freezing) steam-distilled and solvent-extracted using 10 mL of pentane in a condenser-cooled Likens-Nickerson apparatus [17]. Monoterpenes were identified by comparison of retention times and mass spectra to those of authentic standards in gas chromatography with mass spectrometry detection. Quantification was achieved by gas chromatography

with flame ionization detection [17] based upon calibration curves with known amounts of authentic standards and normalization to the peak area of camphor as internal standard.

### **3.3.3 Morphometric measurements.**

The volume of the secretory cells and subcuticular cavity of peppermint secretory cells, as well as the volume densities of subcellular compartments within secretory cells, were estimated based on the morphometric and stereological approaches outlined in [30, 31]. A detailed description of measurements, assumptions and calculations are provided in Supporting Information.

### **3.3.4 Kinetic modeling and simulation.**

A kinetic mathematical model was developed to simulate monoterpene profiles based upon prior knowledge of enzyme expression patterns and kinetic properties. Detailed descriptions of our modeling approaches and the MATLAB source code, in compliance with the MIRIAM guidelines for annotating biochemical models [38], are provided in Supporting Information.

### **3.3.5 Cell-free assaying of recombinant (+)-pulegone reductase activity and inhibition experiments.**

*E coli* BL21(DE3) cells (Invitrogen) were individually transformed with the pSBET plasmids containing peppermint (+)-pulegone reductase (PR) cDNA clone ml579 (AY300163). Transformed *E coli* cells were grown, recombinant protein production induced, cells harvested, and recombinant protein extracted and partially purified



according to [15]. Routine enzyme assays contained 100  $\mu\text{M}$  (+)-pulegone, 500  $\mu\text{M}$  NADPH and 9.2  $\mu\text{g}$  of total protein in 100  $\mu\text{L}$  of 50 mM MOPSO (pH 6.6). Reaction times were adjusted to ensure that no more than 20 % of the available substrate was consumed. Enzymatic reactions were terminated by vortexing with 0.5 mL of pentane and an aliquot of the organic extract was analyzed by GC-FID as above. Kinetic parameters were determined by varying substrate concentration while maintaining other reactants at saturation. Kinetic constants ( $K_m$  and  $V_{max}$ ) were calculated by nonlinear regression analysis (Origin 6.0, OriginLab, Inc.). Substrate inhibition was evaluated in triplicate assays using 15 different (+)-pulegone concentrations between 10 and 800  $\mu\text{M}$ . Preliminary assays to test inhibitory effects on PR activity were performed using varying amounts of (+)-pulegone and (+)-menthofuran (15 different concentrations between 0 and 800  $\mu\text{M}$ ). Triplicate assays were then performed with 0, 20, 60 and 100  $\mu\text{M}$  (+)-pulegone and 0, 80, 160 and 400  $\mu\text{M}$  (+)-menthofuran. Based on these data the mechanism of inhibition was assessed graphically using a Lineweaver-Burk plot [34]. The inhibition constant ( $K_i$ ) for (+)-menthofuran was determined using the Dixon method [35] and nonlinear regression analysis [36].

## **Acknowledgements**

We would like to thank Julia Gothard-Szamosfalvi and Greg Whitney for growing plants. The authors would also like to thank Dr. Ed Davis for valuable discussions and experimental advice, and Iris Lange for technical assistance. R.R.E. thanks the Fulbright program and the University of Antioquia (Medellin, Colombia) for scholarships. This

work was supported in part by the Agricultural Research Center (grants to BML and RBC) and the Mint Industry Research Council (grant to RBC).

### 3.4 References

- [1] Lawrence BM (1981) Essential Oils. Allured Publishing Co., Wheaton IL, pp. 1-81.
- [2] Gershenzon J, Maffei M, Croteau R (1989) Biochemical and histochemical localization of monoterpene biosynthesis in the glandular trichomes of spearmint (*Mentha spicata*). *Plant Physiol* 89:1351-1357.
- [3] McCaskill D, Gershenzon J, Croteau R (1992) Morphology and monoterpene biosynthetic capabilities of secretory cell clusters isolated from glandular trichomes of peppermint (*Mentha × piperita* L.). *Planta* 187:445–454.
- [4] Amelunxen F, Wahlig T, Arbeiter H (1969) Detection of essential oil in isolated glandular trichomes of *Mentha piperita* L (Translated from German). *Pflanzenphysiol* 61:68–72.
- [5] Amelunxen F (1965) Electron microscopy analysis of glandular trichomes of *Mentha piperita* L (Translated from German). *Planta Med* 13:457-473.
- [6] Gershenzon J, McCaskill D, Rajaonarivony JIM, Mihaliak C, Karp F, Croteau R (1992) Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. *Anal Biochem* 200:130–138.
- [7] Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R (2000) Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc Natl Acad Sci USA* 97:2934-2939.
- [8] Croteau R, Davis EM, Ringer KL, Wildung MR (2005) (-)-Menthol biosynthesis and molecular genetics. *Naturwiss* 92:562-577.
- [9] McCaskill D, Croteau R (1995) Monoterpene and sesquiterpene biosynthesis in glandular trichomes of peppermint (*Mentha x piperita*) rely exclusively on plastid-derived isopentenyl diphosphate. *Planta* 197:49–56.
- [10] Eisenreich W, Sagner S, Zenk MH, Bacher A (1997) Monoterpenoid essential oils are not of mevalonoid origin. *Tetrahedron Lett* 38:3889-3892.

- [11] Alonso WR, Rajaonarivony JIM, Gershenzon J, Croteau R (1992) Purification of 4S-limonene synthase, a monoterpene cyclase from the glandular trichomes of peppermint (*Mentha × piperita*) and spearmint (*M. spicata*). *J Biol Chem* 267:7582-7587.
- [12] Colby SM, Alonso WR, Katahira E, McGarvey DJ, Croteau R (1993) 4S-Limonene synthase from the oil glands of spearmint (*Mentha spicata*): cDNA isolation, characterization and bacterial expression of the catalytically active monoterpene cyclase. *J Biol Chem* 268:23016–23024.
- [13] Turner G, Gershenzon J, Nielson EE, Froehlich JE, Croteau R (1999) Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. *Plant Physiol* 120:879-886.
- [14] Lupien S, Karp F, Wildung MR, Croteau R (1999) Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (–)-4S-limonene-3-hydroxylase and (–)-4S-limonene-6-hydroxylase. *Arch Biochem Biophys* 368:181-192.
- [15] Ringer KL, Davis EM, Croteau R (2005) Monoterpene metabolism: cloning, expression and characterization of (–)-isopiperitenol/(–)-carveol dehydrogenase from peppermint and spearmint. *Plant Physiol* 137:863-872.
- [16] Turner GW, Croteau R (2004) Organization of monoterpene biosynthesis in *Mentha*: immunocytochemical localization of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. *Plant Physiol* 136:4215-4227.
- [17] Ringer KL, McConkey ME, Davis EM, Rushing GW, Croteau R (2003) Monoterpene double-bond reductases of the (–)-menthol biosynthetic pathway: isolation and characterization of cDNAs encoding (–)-isopiperitenone reductase and (+)-pulegone reductase of peppermint. *Arch Biochem Biophys* 418:80-92.
- [18] Kjonaas RB, Venkatachalam KV, Croteau R (1985) Metabolism of monoterpenes: oxidation of isopiperitenol to isopiperitenone, and subsequent isomerization to piperitenone, by soluble enzyme preparations from peppermint (*Mentha piperita*) leaves. *Arch Biochem Biophys* 238:49-60.
- [19] Davis EM, Ringer KL, McConkey ME, Croteau R (2005) Monoterpene metabolism: cloning, expression and characterization of menthone reductases from peppermint. *Plant Physiol* 137:873-881.
- [20] Berteau CM, Schalk M, Karp F, Maffei M, Croteau R (2001) Demonstration that menthofuran synthase of mint (*Mentha*) is a cytochrome P450 monooxygenase: cloning, functional expression and characterization of the responsible gene. *Arch Biochem Biophys* 390:279-286.

- [21] Gershenzon J, McConkey ME, Croteau R (2000) Regulation of monoterpene accumulation in leaves of peppermint (*Mentha x piperita* L.). *Plant Physiol* 122:205-214.
- [22] McConkey ME, Gershenzon J, Croteau R (2000) Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint (*Mentha x piperita* L.). *Plant Physiol* 122:215- 233.
- [23] Guenther E (1972) The Essential Oils, Vol. I-VI, Robert E. Krieger Publishing, Huntington, NY.
- [24] Court WA, Roy RC, Pocs R (1993) Effect of harvest date on the yield and quality of the essential oil of peppermint. *Can J Plant Sci* 73:815-824.
- [25] Burbott AJ, Loomis WD (1967) Effects of light and temperature on the monoterpenes of peppermint. *Plant Physiol* 42:20-28.
- [26] Clark RJ, Menary RC (1980) Environmental effects on peppermint, I. Effect of day length, photon flux density, night temperature and day temperature on the yield and composition of peppermint oil. *Aust J Plant Physiol* 7:685-692.
- [27] Mahmoud SS, Croteau RB (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc Natl Acad Sci USA* 98:8915-8920.
- [28] Mahmoud SS, Croteau RB (2003) Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpene reductase. *Proc Natl Acad Sci USA* 100:14481-14486.
- [29] Colson M, Pupier R, Perrin A (1993) Biomathematical analysis of the number of glandular trichomes on leaves of *Mentha x piperita* (Translated from French). *Can J Bot* 71:1202-1211.
- [30] Turner GW, Gershenzon J, Croteau R (2000) Distribution of peltate glandular trichomes on developing leaves of peppermint (*Mentha x piperita* L.). *Plant Physiol* 124:655-664.
- [31] Weibel ER (1979) Stereological Methods, Vol 1, Practical Methods for Biological Morphometry, Weibel ER, ed., Academic Press, London, pp. 63-100.
- [32] Karp F, Mihaliak CA, Harris JL, Croteau R (1990) Monoterpene biosynthesis: specificity of the hydroxylations of (-)-limonene by enzyme preparations from peppermint (*Mentha piperita*), spearmint (*Mentha spicata*), and perilla (*Perilla frutescens*) leaves. *Arch Biochem Biophys* 276:219-226.

- [33] Lawrence BM (2006) *Mint: The Genus Mentha*. CRC Press, Boca Raton, FL, 556 pages.
- [34] Lineweaver H, Burk D (1934) The determination of enzyme dissociation constants. *J Am Chem Soc* 56:658—666.
- [35] Dixon M (1952) The determination of enzyme inhibitor constants. *Biochem J* 55:170-171.
- [36] Kakkar T, Boxenbaum H, Mayersohn M (1999) Estimation of  $K_i$  in a competitive enzyme-inhibition model: comparisons among three methods of data analysis. *Drug Metab Dispos* 27:756-762.
- [37] Gershenzon J, McCaskill D, Rajaonarivony JIM, Mihaliak C, Karp F, Croteau R (1992) Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. *Anal Biochem* 200:130-138.
- [38] Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P et al. (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 23:1509-1515.

### 3.5 Supplementary Material

#### SI Appendix<sup>1</sup>

##### 1. Estimating enzyme and metabolite concentrations in individual glandular trichomes

Peppermint glandular trichomes harbor three different cells types (outlined in Scheme 1). Among these cell types the secretory cells (eight-celled disk highlighted in Scheme 1) are responsible for the biosynthesis of monoterpenes in peppermint. We inferred the number of active secretory cells from published data for greenhouse-grown plants [1] and growth chamber-grown plants [2], estimated the volumes of the secretory cell disk and the monoterpenoid essential oil storage cavity, and used these values to calculate enzyme and metabolite concentrations in individual glandular trichomes.

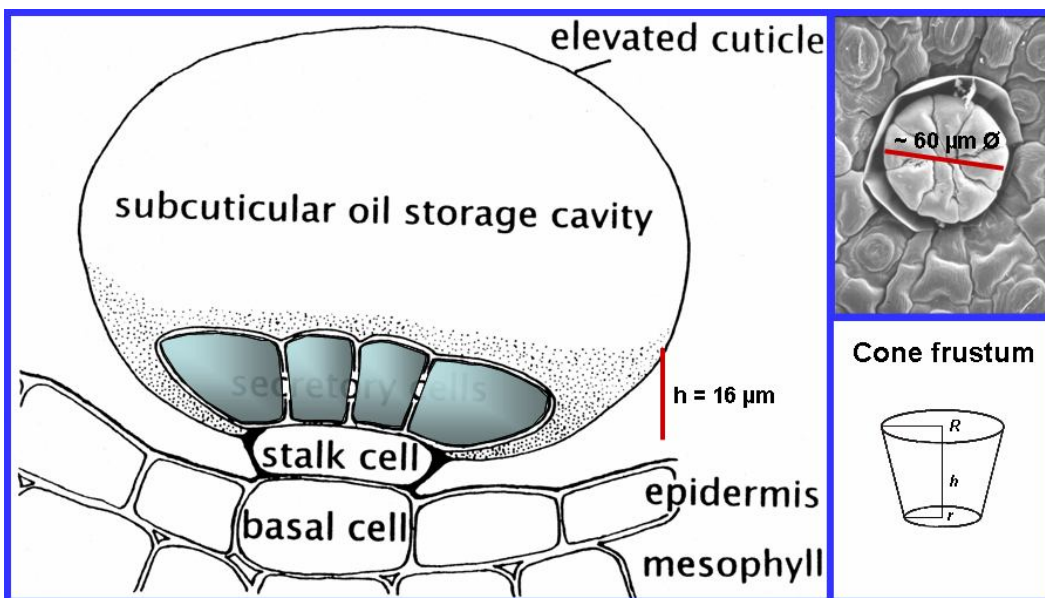
##### Estimating the number of glandular trichomes active in monoterpene biosynthesis

Leaf age [d]	8	12	15	21	24
Number of active glandular trichomes (GH)	3,000	6,000	15,000	15,000	14,000
Number of active glandular trichomes (GC)	1,700	4,600	7,500	7,500	6,500

## Estimating the volume of the secretory cell disk of glandular trichomes

Based on microscopy-based measurements, the secretory cell cluster was determined to have an average diameter of 60  $\mu\text{m}$ , with a height of 16  $\mu\text{m}$  (SI Scheme 1). If the shape of the secretory cell cluster is approximated by a frustum of a cone (SI Scheme 1), the volume can be calculated as  $\frac{1}{3} \pi h (R^2 + R r + r^2) = 2.35 \cdot 10^{-5} \mu\text{L}$ .

### Scheme 1



## Estimating the volume densities of subcellular compartments

The volumes of the subcellular compartments in which the monoterpene metabolic enzymes are located were determined using morphometric and stereological approaches:

The volume of an individual secretory phase gland cell was estimated from measured dimensions (for details see above) to be  $1/8 (2.35 \times 10^{-5} \mu\text{l}) = 2.9 \times 10^{-6} \mu\text{l}$ . The volumes of organelles within a secretory cell were estimated by two different approaches using ImageJ software (<http://rsb.info.nih.gov/ij>). The percentage of micrograph area (volume density) covered by leucoplasts and mitochondria was directly calculated by encircling organelles with a calibrated measuring tool. Such direct measurements using the ImageJ drawing tool were feasible for larger organelles, but were impractical for the tubular smooth ER, which consists of numerous interconnected, narrow diameter tubes. The volume densities of plastids, mitochondria, ER, vacuoles and cytosol were also determined by randomly superimposing a stereological grid overlay on gland cell micrographs and by counting the number of intercepts these organelles made with test points. The grid overlay plug-in for ImageJ was obtained at <http://rsb.info.nih.gov/ij/plugins/grid.html>. The test points consisted of the intersections of horizontal and vertical lines (the corners of grid squares) separated by spacing representing  $1 \mu\text{m}$  on the micrographs.

Representative micrographs of secretory cells from six different secretory-phase peltate glandular trichomes were used for the morphometric measurements. These were chosen to include both apical and basal regions of gland cells, since there is some polarity in the distribution of organelles [3]. All specimens were preserved by high-pressure freezing and freeze-substitution in order to ensure good preservation of gland cell ultrastructure. However, this method leads to an extraction of low molecular weight lipids during the freeze-substitution process, so that stereological estimates of volume density for vacuolar and cytoplasmic monoterpene droplets could not be obtained.



The directly determined volumes for plastids and mitochondria were very similar to those obtained with volume densities calculations from test-point counts, confirming the quality of our stereological estimates. The average volume density for leucoplasts was 13.3 % area (directly measured) and 13.9 % (stereological test-point intercepts). The difference between these values (0.6 %) is smaller than the standard deviation between leucoplast volume densities for the six individual glands ( $\sigma_{vd} = 3.25$  directly measured;  $\sigma_{vd} = 4.02$  when determined stereologically). The difference between the estimates for the mitochondrial volumes based on these methods (1.0 %) is slightly larger than the standard deviations between individual glands ( $\sigma_{vd} = 0.6$  directly measured;  $\sigma_{vd} = 0.8$  when determined stereologically).

<b>Organelles</b>	<b>% Cross-sectional area, volume density</b>	<b><math>\sigma_{vd}</math></b>	<b>Estimated volume per gland cell, <math>\mu\text{l}</math></b>	<b>Average diameter, <math>\mu\text{m}</math></b>	<b><math>\sigma_{di}</math></b>	<b>Surface area per gland cell, <math>\mu\text{m}^2</math></b>
Leucoplasts	13.9	4.02	$0.41 \times 10^{-6}$	n.m.		---
Mitochondria	4.4	0.6	$0.13 \times 10^{-6}$	0.47 (n=95)	0.11	$1.66 \times 10^3$
ER	36.5	3.5	$1.07 \times 10^{-6}$	0.07 (n=265)	0.04	$1.53 \times 10^4$
Vacuoles	16.2	4.3	$0.48 \times 10^{-6}$	n.m.		---
Cytoplasm	20.4	3.1	$0.60 \times 10^{-6}$	n.m.		---
Other	8.6					---

The number of point counts required to obtain accurate volume densities were calculated according to published methods [4]. The required number of point counts ( $P_c$ ) is proportional to the standard error and is reduced with larger organelle volume density ( $V_{va}$ ) and with a larger number of replicate specimens ( $m$ ) by the equation:

$$P_c = (t_{\alpha}^2/m*d^2)(1-V_{va}/V_{va}), \quad \text{where } d \text{ is the confidence interval.}$$

For example, the required number of point counts (per gland) to obtain a 95 % probability for accuracy within a 10 % confidence interval would be 409.5 (total of 2457 counts) for leucoplasts. Our actual count consisted of 2266 test-point counts with an average of 378 counts per specimen (with a range of 233 to 476), which results in estimated confidence intervals at the 95% probability level of 10.4 % for leucoplasts, 17.6 % for mitochondria, 9.6 % for vacuoles, and 5.5 % for ER. We also calculated the total surface area of mitochondria and ER within secretory cells by assuming that mitochondria are spherical and the smooth ER consists of narrow cylinders.

Based on the available literature the subcellular distribution of enzymes in peppermint secretory cells is as follows:

<b>Enzyme</b>	<b>Compartment</b>	<b>Reference</b>
(-)-Limonene synthase	Leucoplasts	[5]
(-)-Limonene 3-hydroxylase	Endoplasmic reticulum	[6]
(-)- <i>trans</i> -Isopiperitenol dehydrogenase	Mitochondria	[7]
(-)-Isopiperitenone reductase	Cytosol	Turner et al, unpublished
(+)- <i>cis</i> -Isopulegone isomerase	Cytosol	[8]
(+)-Pulegone reductase	Cytosol	[7]
(-)-Menthone:(-)-menthol reductase	Cytosol	Turner et al, unpublished
(-)-Menthone:(+)-neomenthol reductase	Cytosol	Turner et al, unpublished
(+)-Menthofuran synthase	Endoplasmic reticulum	[9]

### **Estimating the monoterpene amount synthesized in a glandular trichome**

We used two independent methods to calculate the amounts of monoterpenes produced per individual glandular trichome:

1.) Monoterpene amounts obtained from GC-FID analyses of steam distillations were divided by the number of oil-bearing glandular trichomes:

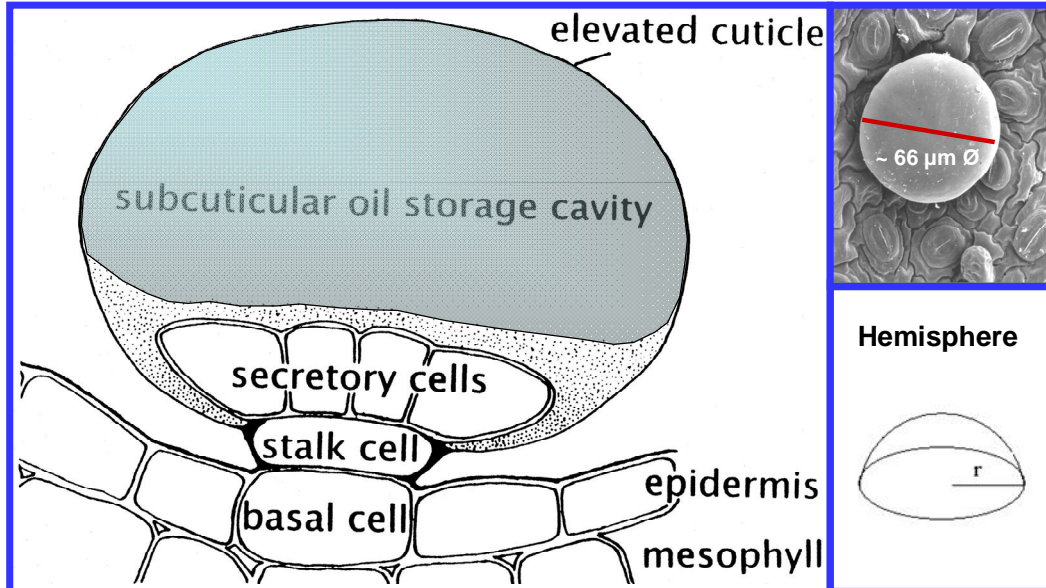
Leaf age [d]	8	12	15	21	24
Number of active glandular trichomes (greenhouse)	3,000	6,000	15,000	15,000	14,000
Determined monoterpene amount [ $\mu\text{g}$ per leaf]	98	267	1,162	1,271	1,208
Calculated monoterpene amount [ng per gland]	33	45	77	85	86
Calculated monoterpene amount [pmol per gland]	220	300	513	566	572

Leaf age [d]	8	12	15	21	24
Number of active glandular trichomes (GC)	1,700	4,600	7,500	7,500	6,500
Determined monoterpene amount [ $\mu\text{g}$ per leaf]	43	150	629	530	482
Calculated monoterpene amount [ng per gland]	26	33	84	71	74
Calculated monoterpene amount [pmol per gland]	173	220	559	473	493

The gradient from 8 to 24 d reflects the fact that the number of mature glandular trichomes (with maximum essential oil volume) increases over time.

2.) The volume of the essential oil-filled subcuticular cavity of mature glandular trichomes (area filled with light blue color in SI Scheme 2) was calculated based on the approximation of its shape as a hemisphere ( $\frac{2}{3} \pi r^3$ ) with a base diameter  $66 \mu\text{m}$  (determined using microscopic measurements), from which the volume of a cone frustum (determined as outlined above) was subtracted. The estimated volume ( $7.53 \cdot 10^{-5} \mu\text{L}$  for a mature glandular trichome) was then multiplied by the known essential oil density ( $0.9 \text{ mg} / \mu\text{L}$ ) to obtain a monoterpene amount of  $68 \text{ ng}$  per gland. An average monoterpene molecular mass of  $150 \text{ g} / \text{mol}$  was assumed for converting these values into molar amounts per gland. The calculated value ( $453 \text{ pmol}$  per gland) is very similar to that estimated from GC data and glandular trichome counts at maturity ( $572$  and  $493 \text{ pmol}$  per gland for greenhouse and growth chamber-grown plants, respectively).

## Scheme 2

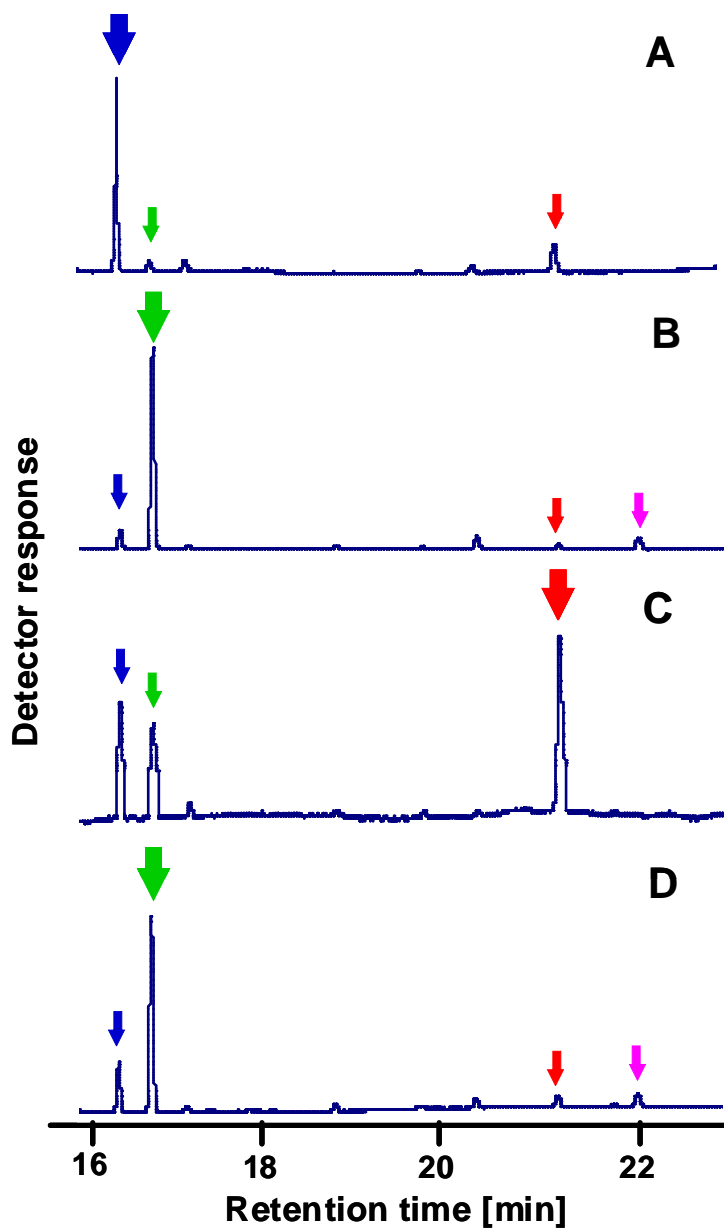


### Measurement of monoterpene profiles in isolated peppermint secretory cells

Peppermint secretory cells were isolated based on a published protocol [10] and steam-distilled and solvent-extracted using 10 mL of pentane in a condenser-cooled Likens-Nickerson apparatus [11]. Monoterpenes were identified by comparison of retention times and mass spectra to those of authentic standards in gas chromatography with mass spectrometry detection. Quantification was achieved by gas chromatography with flame ionization detection [11] based upon calibration curves with known amounts of authentic standards and normalization to the peak area of camphor as internal standard. SI Scheme 3 shows a GC-FID-based monoterpene analysis of peppermint secretory cell clusters isolated from 20-day-old plants grown in the greenhouse (**A**), 20-day-old plants grown under low light intensity (**B**), 50-day-old plants grown in the greenhouse (**C**), and 50-day-old plants grown under low light intensity (**D**). Arrows indicate the retention times of the

major monoterpenes: (-)-menthone, blue; (+)-menthofuran, green; (-)-menthol, red; (+)-pulegone, purple. It should be noted that the (+)-menthofuran concentration was 50 times higher in sample (D) than in sample (C) (20 mM versus 400  $\mu$ M).

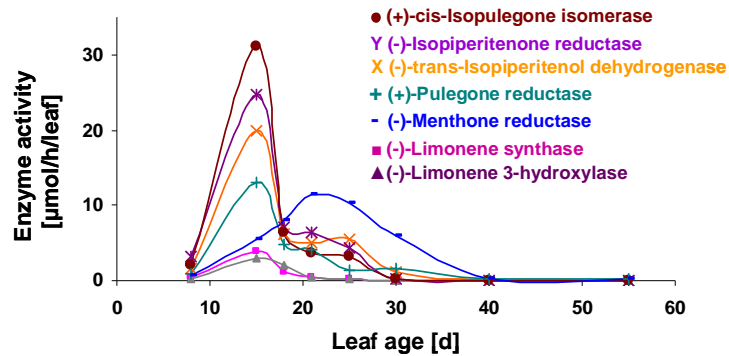
Scheme 3



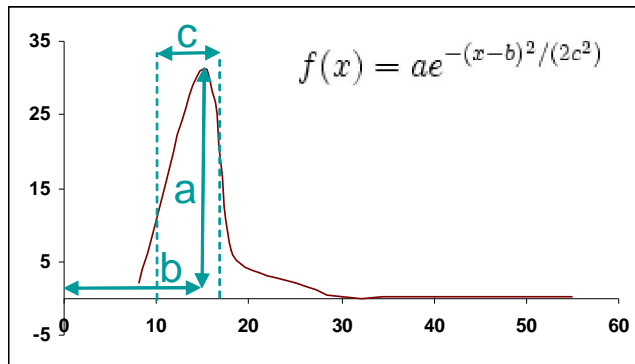
## 2. Developmental patterns of monoterpene biosynthetic enzyme activities

For the majority of currently available kinetic models it is assumed that the amounts of biosynthetic enzymes remain constant for the duration of the experimental period. In peppermint glandular trichomes, the biosynthesis of monoterpenes involves dynamic changes in the activities of biosynthetic enzymes [12]. We used the available experimental data on developmental changes in biosynthetic enzyme activities to approximate changes in enzyme amounts with a Gaussian function:

### A) Experimental enzyme activity data



### B) Example of a Gaussian function to approximate enzyme activity data



The Gaussian function represents the following variables:

$$f(t) = \text{Comp} * \text{Enz} * \exp(-((t-b)^2)/(2*c^2))$$

where Comp = Factor to adjust for the volume density of the compartment in which a particular enzyme is active

Enz = Enzyme concentration in the compartment in which it resides [uM]

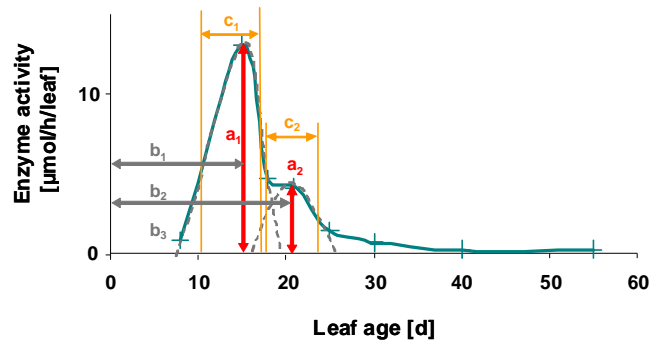
t = Time [s]

b = Factor defining the position of the center of the Gaussian peak for enzyme activity [s]

c = Factor defining the width of the Gaussian peak for enzyme activity at half maximum [s]

### C) Example of the use of two Gauss functions to approximate enzyme activity data

To approximate the curve for developmental patterns of enzyme activities with non-Gaussian shapes we used more than one Gaussian function. The pattern of (+)-pulegone reductase activity will serve as an example (plot of Gaussian graph with gray dotted lines):





### 3. Generating a system of ordinary differential equations to describe kinetic properties of enzymes

The Michaelis-Menten rate equation, as developed by Briggs and Haldane [13], allows calculating the change of the concentration of a metabolite based on the rate of enzymatic formation and turnover. Using the monoterpene pathway intermediate (-)-limonene as an example we obtain the following:

$$\frac{d[IPPol]}{dt} = \frac{k_{c2}E2[LIM]}{[LIM] + KM2} - \frac{k_{c3}E3[IPPol]}{[IPPol] + KM3}$$

with  $K_{c2}E2$ ,  $K_{cat}(\text{limonene 3-hydroxylase})$ ;  $K_{c3}E3$ ,  $K_{cat}(\text{isopiperitenol dehydrogenase})$ ;  $IPPol$ , (-)-isopiperitenol;  $LIM$ , (-)-limonene;  $KM2$ ,  $K_m(\text{limonene 3-hydroxylase})$ ;  $KM3$ ,  $K_m(\text{isopiperitenol dehydrogenase})$ .

The same formalism was also used for the mathematical description of the properties of the other enzymes in the monoterpene biosynthetic pathway. A special case is the reaction catalyzed by (-)-menthone:(+)-neomenthol reductase, which is a reversible enzyme, and the expression for (-)-menthone turnover needed to be adjusted accordingly.

The reaction catalyzed by (+)-pulegone reductase yields (-)-menthone and (+)-isomenthone in a 10 : 1 ratio. We used two separate expressions for these reactions (basically treating the two reactions as being catalyzed by two different enzymes).

The enzyme (-)-menthone:(-)-menthol reductase accepts two substrates ((-)-menthone and (+)-isomenthone) and converts them into two different products ((-)-menthol and (+)-

neoisomenthol). Since the mechanism of this reaction is unknown, these two reactions are treated as being catalyzed by two different enzymes. The same is true for the enzyme (-)-menthone:(+)-neomenthol reductase (substrates: (-)-menthone and (+)-isomenthone; products: (+)-neomenthol and (+)-isomenthol).

We also tested several hypotheses with regard to the effect of potential feedback inhibition. For example, we considered the effect of competitive, uncompetitive and noncompetitive inhibition of (+)-pulegone reductase by (+)-menthofuran using modified Michaelis-Menten rate equations (the effect on (+)-pulegone is shown as an example):

1.) Competitive inhibition ( $K_{ic}$  = inhibition constant for competitive mechanism)

$$\frac{d[PUL]}{dt} = \frac{kc5E5[CIPUL]}{[CIPUL] + KM5} - \frac{kc7aE7a[PUL]}{KM7a \left(1 + \frac{[MF]}{kic}\right) + [PUL]} - \frac{kc7bE7b[PUL]}{KM7b \left(1 + \frac{[MF]}{Kic}\right) + [PUL]} - \frac{kc6E6[PUL]}{[PUL] + KM6}$$

2.) Uncompetitive inhibition ( $K_{iu}$  = inhibition constant for uncompetitive mechanism)

$$\frac{d[PUL]}{dt} = \frac{kc5E5[CIPUL]}{[CIPUL] + KM5} - \frac{kc7aE7a[PUL]}{[PUL] \left(1 + \frac{[MF]}{kiu}\right) + KM7a} - \frac{kc7bE7b[PUL]}{[PUL] \left(1 + \frac{[MF]}{kiu}\right) + KM7b} - \frac{kc6E6[PUL]}{[PUL] + KM6}$$

3.) Noncompetitive inhibition ( $K_{in}$  = inhibition constant for noncompetitive mechanism)

$$\frac{d[PUL]}{dt} = \frac{kc5E5[CIPUL]}{[CIPUL]+KM5} - \frac{kc7aE7a[PUL]}{(KM7a+[PUL])\left(1+\frac{[MF]}{kin}\right)} - \frac{kc7bE7b[PUL]}{(KM7b+[PUL])\left(1+\frac{[MF]}{kin}\right)} - \frac{kc6E6[PUL]}{[PUL]+KM6}$$

In addition to assessing the inhibition of (+)-pulegone reductase by (+)-menthofuran, we also considered substrate inhibition of (+)-pulegone reductase (as demonstrated in the present paper; the example below assumes a competitive inhibition mechanism):

$$\begin{aligned} \frac{d[PUL]}{dt} = & \frac{kc5E5[CIPUL]}{[CIPUL]+KM5} - p^* \frac{kc7aE7a[PUL]}{KM7a\left(1+\frac{z[M]}{kic}\right)+[PUL]} - p^* \frac{kc7bE7b[PUL]}{KM7b\left(1+\frac{z[M]}{Kic}\right)+[PUL]} \\ & - q^* \frac{kc7bE7b[PUL]}{KM7b+\left(1+\frac{[PUL]}{Kis}\right)[PUL]} - q^* \frac{kc7aE7a[PUL]}{KM7a+\left(1+\frac{[PUL]}{Kis}\right)+[PUL]} - \frac{kc6E6[PUL]}{[PUL]+KM6} \end{aligned}$$

where  $p$  and  $q$  are factors that determine which percentage of (+)-pulegone reductase is affected by competitive inhibition and substrate inhibition.

The following kinetic parameters were used in the initial mathematical model:

Enzyme	Km [μM]	Kcat [s <sup>-1</sup> ]	Reference
(-)-Limonene synthase	20	0.3	[14]
(-)-Limonene 3-hydroxylase	18	1*	[15]
(-)- <i>trans</i> -Isopiperitenol dehydrogenase	72	0.002	[16]
(-)-Isopiperitenone reductase	1.0	1.3	[11]
(+)- <i>cis</i> -Isopulegone isomerase	270	?*	[8]
(+)-Pulegone reductase	2.3	1.8	[11]
(-)-Menthone:(-)-menthol reductase	3.0	0.6	[17]
with (+)-isomenthone as substrate	41	?*	[17]
(-)-Menthone:(+)-neomenthol reductase	674	0.06	[17]
with (+)-isomenthone as substrate	1000	?*	[17]
(+)-Menthofuran synthase	10*	0.9*	[9]

\* These values could not be obtained from the literature and have thus been estimated. All other kinetic values were reported in the papers cited above.

## 4.1 Mint monoterpene metabolism in plants cultivated under greenhouse conditions

(first generation model)

### *Mathematical Model of Peppermint Monoterpene Biosynthesis I (for greenhouse-grown plants)*

```
function xdot = mm5(t,x,flag,E)

% This function calculates monoterpene amounts (58 day time course) in
% leaves of peppermint plants grown in a greenhouse with supplemental
% lighting from sodium vapor lights. A mechanism following regular
% Michaelis-Menten-type kinetics is assumed for all enzymes.

% The model also takes into account that each enzyme shows a particular
% transient pattern of expression. This pattern is approximated by a
% Gaussian function.

% Metabolite Nomenclature

%[GPP]=x(1)          Geranyl diphosphate
%[LIM]=x(2)          (-)-Limonene
%[IPPol]=x(3)        (-)-trans-Isopiperitenol
%[IPPone]=x(4)       (-)-Isopiperitenone
%[CIPUL]=x(5)        (+)-cis-Isopulegone
%[PUL]=x(6)          (+)-Pulegone
%[MF]=x(7)           (+)-Menthofuran
%[IMone]=x(8)        (+)-Isomenthone
%[Mone]=x(9)         (-)-Menthone
%[NMol]=x(10)        (+)-Neomenthol
%[Mol]=x(11)         (-)-Menthol
%[IMol]=x(12)        (+)-Isomenthol
%[NIMol]=x(13)       (+)-Neoisomenthol

% Kinetic Parameters

% kc units: [1/s]    kc = Kcat
% KM units: [uM]
% Ki units: [uM]

kc1 = 0.3;          %(-)-Limonene synthase (LS)
KM1 = 20;           %
kc2 = 1;            %(-)-Limonene 3-hydroxylase (L3H)
KM2 = 18;           %
kc3 = 0.002;       %(-)-trans-Isopiperitenol dehydrogenase (IsoDH)
KM3 = 72;           %
kc4 = 1.3;         %(-)-Isopiperitenone reductase (IsoR)
KM4 = 1;           %
kc5 = 1;           %(+)-cis-Isopulegone isomerase (IsoI)
KM5 = 270;         %
kc6 = 0.9;         %(+)-Menthofuran synthase (MFS)
KM6 = 10;          %
```

```

kc7a = 1.8;      %(+)-Pulegone reductase (PR; product: (-)-menthone)
KM7a = 2.3;      %
kc7b = 1.8;      %(+)-Pulegone reductase (PR; product:
                (+)-isomenthone)
KM7b = 2.3;      %
kc8a = 0.6;      %(-)-Menthone:(-)-menthol reductase (MMR; substrate:
                (-)-menthone)
KM8a = 3;        %
kc8b = 0.6;      %(-)-Menthone:(-)-menthol reductase (MMR; substrate:
                (+)-isomenthone)
KM8b = 41;       %
kcf9a = 0.06;    %(-)-Menthone:(+)-neomenthol reductase (MNR;
                substrate: (-)-menthone); forward reaction)
KMf9a = 674;     %
kcb9a = 0.06;    %(-)-Menthone:(+)-neomenthol reductase (MNR;
                substrate: (-)-menthone); backward reaction)
KMb9a = 1200;    % estimated
kc9b = 0.06;     %(-)-Menthone:(+)-neomenthol reductase (MNR;
                substrate: (+)-isomenthone)
KM9b = 1000;

%Transient Patterns of Enzyme Activity (first peak)

%Enzyme activity patterns are approximated by a Gauss function:

%f(t) = Comp * a * exp((-t-b).^2)/(2*(c)^2))

%where  Comp = Factor to adjust for the volume density of the
compartment in which a particular enzyme is active [Dimensionless]

%      a      = Factor defining the height of the Gaussian peak for
enzyme activity [ Units of concentration ]
%      t      = Time [s]
%      b      = Factor defining the position of the center of the
Gaussian peak for enzyme activity [s]
%      c      = Factor defining the width of the Gaussian peak for
enzyme activity at half maximum [s]

b1=1296000; % Defines the position of the center of the Gaussian
peak for enzyme activity. Relevant to the following
enzyme activities: LS, L3H, IsoDH, IsoR, IsoI, MFS, PR

c1=800000; % Defines the width of the Gaussian peak for enzyme
activity at half maximum. Relevant to the following
enzyme activities: LS,L3H, IsoDH, IsoR, IsoI, MFS, PR

b5=2600000; % Defines the position of the center of the Gaussian
peak for enzyme activity. Relevant to the following
enzyme activities: MMR, MNR

c5=900000; % Defines the width of the Gaussian peak for enzyme
activity at half maximum. Relevant to the following
enzyme activities: MR, MNR

E1= (0.139)*0.017*exp((-t-b1).^2)/(2*(c1)^2)); %LS
E2= (0.365)*0.003*exp((-t-b1).^2)/(2*(c1)^2)); %L3H
E3= (0.044)*10*exp((-t-b1).^2)/(2*(c1)^2)); %IsoDH

```

```

E4= (0.204)*0.34*exp((-t-b1).^2)/((2*c1)^2); %IsoR
E5= (0.204)*0.34*exp((-t-b1).^2)/((2*c1)^2); %IsoI
E6= (0.365)*0.00015*exp((-t-b1).^2)/(2*(c1)^2); %MFS
E7a=(0.204)*0.0012*exp((-t-b1).^2)/(2*(c1)^2); %PR (product:
(-)-menthone)
E7b=(0.204)*0.00012*exp((-t-b1).^2)/(2*(c1)^2); %PR (product:
(+)-isomenthone)
E8a=(0.204)*0.0014*exp((-t-b5).^2)/(2*(c5)^2); %MMR (substrate:
(-)-menthone)
E8b=(0.204)*0.0014*exp((-t-b5).^2)/(2*(c5)^2); %MMR (substrate:
(+)-isomenthone)
E9a=(0.204)*0.001*exp((-t-b5).^2)/(2*(c5)^2); %MNR (substrate:
(-)-menthone)
E9b=(0.204)*0.001*exp((-t-b5).^2)/(2*(c5)^2); %MNR (substrate:
(+)-isomenthone)

%Species Equations

if t< 1555200 % (patterns of enzymes from 0 to 18 days after leaf
initiation)

xdot=[-kc1*E1*x(1)/(x(1)+KM1); % Variation of
GPP
kc1*E1*x(1)/(x(1)+KM1)-kc2*E2*x(2)/(x(2)+KM2); % Variation of
LIM
kc2*E2*x(2)/(x(2)+KM2)-kc3*E3*x(3)/(x(3)+KM3); % Variation of
IPPol
kc3*E3*x(3)/(x(3)+KM3)-kc4*E4*x(4)/(x(4)+KM4); % Variation of
IPPone
kc4*E4*x(4)/(x(4)+KM4)-kc5*E5*x(5)/(x(5)+KM5); % Variation of
CIPUL
kc5*E5*x(5)/(x(5)+KM5)-kc7a*E7a*x(6)/(KM7a+x(6))-
kc7b*E7b*x(6)/(KM7b+x(6))-
kc6*E6*x(6)/(x(6)+KM6); % Variation of
PUL
kc6*E6*x(6)/(x(6)+KM6); % Variation of
MF
kc7b*E7b*x(6)/(KM7b+x(6))-kc8b*E8b*x(8)/(x(8)+KM8b)-
kc9b*E9b*x(8)/(x(8)+KM9b); % Variation of
(+)-IMone
kc7a*E7a*x(6)/(KM7a+x(6)) - ((kcf9a*E9a*x(9)/Kmf9a-
(kcb9a*E9a*x(10)/Kmb9a))/(1+(x(9)/Kmf9a)+(x(10)/Kmb9a)))-
kc8a*E8a*x(9)/(x(9)+KM8a); % Variation of
Mone
((kcf9a*E9a*x(9)/Kmf9a-
(kcb9a*E9a*x(10)/Kmb9a))/(1+(x(9)/Kmf9a)+(x(10)/Kmb9a))); % Variation of
NMol
kc8a*E8a*x(9)/(x(9)+KM8a); % Variation of
Mol
kc9b*E9b*x(8)/(x(8)+KM9b); % Variation of
IMol
kc8b*E8b*x(8)/(x(8)+KM8b)]; % Variation of
NIMol

else t>= 1555200 % (patterns of enzymes from 21 - 60 days after leaf

```

```

initiation)

%Transient Patterns of Enzyme Activity (second peak)

    b2=1814400;   % Defines the position of the center of the second
Gaussian peak   for enzyme activity.
                % Relevant to the following enzyme activities: PR
    c2=1420000;   % Defines the width of the second Gaussian peak for
enzyme          activity at half maximum.
                % Relevant to the following enzyme activities: PR

    b4=2160000;   % Defines the position of the center of the second
Gaussian peak   for enzyme activity.
                % Relevant to the following enzyme activities: IsoDH,
IsoR, IsoI      % Relevant to the following enzyme activities: IsoDH,
    c4=170000;   % Defines the width of the second Gaussian peak for
enzyme          activity at half maximum.
                % Relevant to the following enzyme activities: IsoDH,
IsoR, IsoI

E3=(0.044)*1*exp((- (t-b4).^2)/(2*(c4)^2));           %IsoDH
E4=(0.204)*0.0044*exp((- (t-b2).^2)/(2*(c2)^2));     %IsoR
E5=(0.204)*0.00044*exp((- (t-b2).^2)/(2*(c2)^2));   %IsoI
E7a=(0.204)*0.0001*exp((- (t-b2).^2)/(2*(c2)^2));   %PR (product: (-)-
menthone)
E7b=(0.204)*0.00001*exp((- (t-b2).^2)/(2*(c2)^2)); %PR (product:
(+) -isomenthone)

%Species Equations

    xdot=[-kc1*E1*x(1)/(x(1)+KM1);                    % Variation of
GPP
          kc1*E1*x(1)/(x(1)+KM1)-kc2*E2*x(2)/(x(2)+KM2); % Variation of
LIM
          kc2*E2*x(2)/(x(2)+KM2)-kc3*E3*x(3)/(x(3)+KM3); % Variation of
IPPol
          kc3*E3*x(3)/(x(3)+KM3)-kc4*E4*x(4)/(x(4)+KM4); % Variation of
IPPone
          kc4*E4*x(4)/(x(4)+KM4)-kc5*E5*x(5)/(x(5)+KM5); % Variation of
CIPUL
          kc5*E5*x(5)/(x(5)+KM5)-kc7a*E7a*x(6)/(KM7a+x(6))-
          kc7b*E7b*x(6)/(KM7b+x(6))-
          kc6*E6*x(6)/(x(6)+KM6);                    % Variation of
PUL
          kc6*E6*x(6)/(x(6)+KM6);                    % Variation of
MF
          kc7b*E7b*x(6)/(KM7b+x(6))-kc8b*E8b*x(8)/(x(8)+KM8b)-
          kc9b*E9b*x(8)/(x(8)+KM9b);
                                                    % Variation of
(+) -IMone
          kc7a*E7a*x(6)/(KM7a+x(6)) - ((kcf9a*E9a*x(9)/KMf9a-
          (kcb9a*E9a*x(10)/KMb9a))/(1+(x(9)/KMf9a)+(x(10)/KMb9a)))-

```

```

        kc8a*E8a*x(9)/(x(9)+KM8a);           % Variation of
Mone
        ((kcf9a*E9a*x(9)/KMf9a-
(kcb9a*E9a*x(10)/KMb9a))/(1+(x(9)/KMf9a)+(x(10)/KMb9a)));
        % Variation of
NMol
        kc8a*E8a*x(9)/(x(9)+KM8a);           % Variation of
Mol
        kc9b*E9b*x(8)/(x(8)+KM9b);           % Variation of
IMol
        kc8b*E8b*x(8)/(x(8)+KM8b)];           % Variation of
NIMol
end

```

Matlab Code for simulation of the remaining environmental conditions can be found online at the PNAS website.



## CHAPTER 4

### Monoterpenoid essential oil yield in peppermint is determined by the density and size distribution of glandular trichomes

Rigoberto Rios-Esteva<sup>a,c</sup>, Iris Lange<sup>a</sup>, James M. Lee<sup>c</sup>, and B. Markus Lange<sup>a,b,c,\*</sup>

<sup>a</sup>*Institute of Biological Chemistry*, <sup>b</sup>*M. J. Murdock Metabolomics Laboratory*, and <sup>c</sup>*School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA 99164-6340*

**Author contributions:** RRE, and BML designed research; RRE and IL performed research; RRE, IL, JML, and BML analyzed data; and RRE, and BML wrote the paper.

**Abbreviations:** DXS, -deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMK, 4, 4-(cytidine 5'-diphospho)-2C-methyl-D-erythritol 4-phosphate kinase; HDS, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate synthase; LS, (-)-limonene synthase; L3H, (-)-limonene 3-hydroxylase; PR, (+)-pulegone reductase; MFS, (+)-menthofuran synthase.

\* Corresponding author. Fax: +1 509 335-3794. *E-mail address:* lange-m@wsu.edu.

Manuscript submitted to: *Archives of Biochemistry and Biophysics*

### Abstract

A reassessment of transgenic plants expressing an antisense version of (+)-menthofuran synthase (MFS) transcript revealed that, in addition to previously reported effects on essential oil quality, the oil yield in one particular line, designated MFS7, was also enhanced (1.35-fold up) compared to wild-type controls. Based on real-time quantitative PCR assays, the monoterpenoid essential oil composition in MFS7 plants, but not increased yield compared to wild-type controls, could be explained by gene expression patterns. A microscopic evaluation of leaf surfaces demonstrated that,

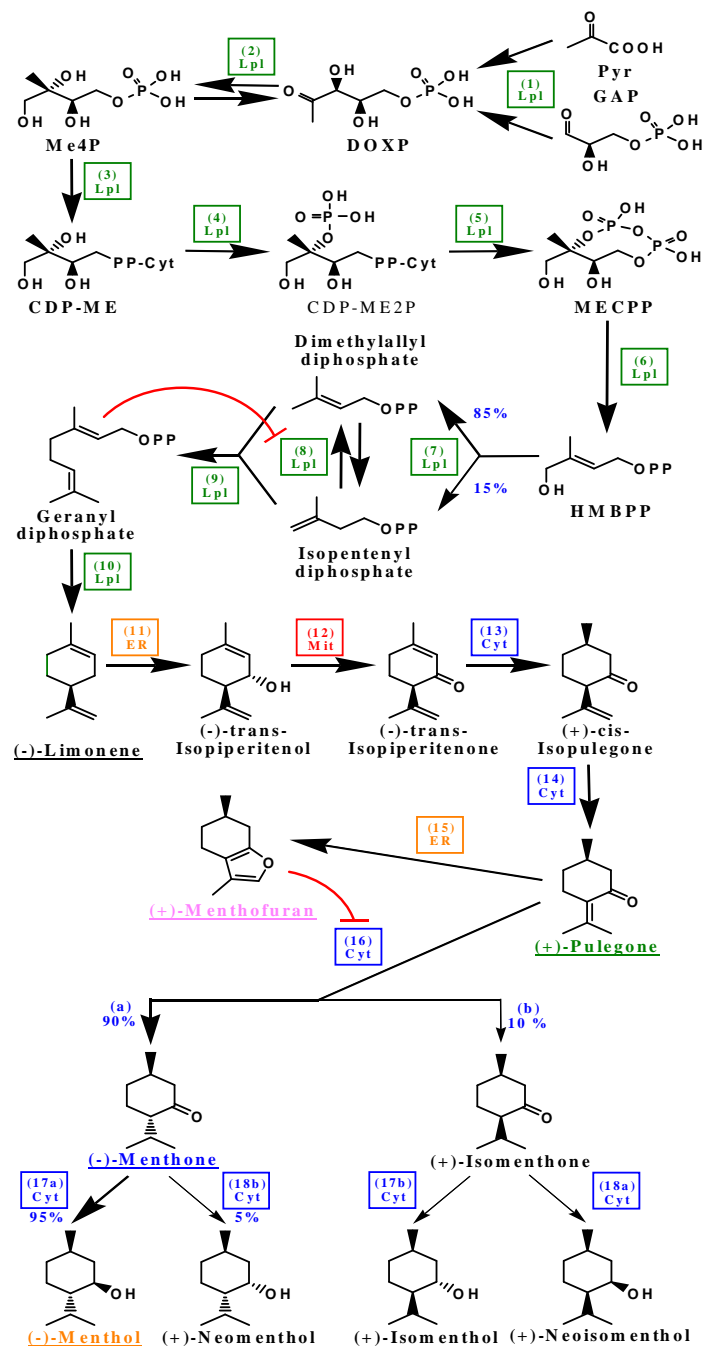
compared to controls, MFS7 plants had a higher density (1.12-fold up) of glandular trichomes, the specialized anatomical structures responsible for the synthesis and storage of essential oils. Across various experiments, glandular trichome density was strongly correlated with monoterpenoid essential oil yields in both MFS7 and wild-type leaves ( $R^2 = 0.994$ ). Based on these experimental data, our existing kinetic mathematical model was updated and simulations shown to match the experimentally observed environmental and genotypic variation of monoterpene yield.

## 4.1 Introduction

The essential oil distilled from peppermint (*Mentha x piperita*) leaves is used in numerous consumer products (e.g., chewing gum, toothpaste, and mouthwash), as a flavor in the confectionary and pharmaceutical industries, and as a source of active ingredients for aromatherapy. Peppermint oil consists primarily of *p*-menthane-type monoterpenes, with smaller amounts of other monoterpenes, and very minor quantities of sesquiterpenes [1]. The essential oil is synthesized and accumulated in specialized anatomical structures called peltate glandular trichomes [2,3]. These trichomes contain secretory cells, arranged in an eight-celled disk, which are responsible for the synthesis of essential oil. Essential oil is secreted into an emerging cavity formed by the separation of a preformed layer of cuticular material [4]. Over the last two decades, peppermint has been developed as a model system for the biosynthesis of monoterpenoid essential oils [5]. Based on the rich body of published data on peppermint monoterpene biosynthesis, we have recently developed a kinetic mathematical model to simulate the variation of

monoterpene profiles under certain environmental conditions [6]. Based on our modeling, we predicted that the monoterpene profiles observed in leaves of plants grown under low light conditions could only be explained if one assumed that (+)-menthofuran, a dead-end side product, acted as a competitive inhibitor against (+)-pulegone, the primary substrate of the branch point enzyme (+)-pulegone reductase (Fig. 4.1). Follow-up biochemical studies established that this prediction was correct [6], thus illustrating the utility of integrating mathematical modeling with experimental testing in an iterative cycle.

To further advance our modeling efforts, we identified a need for assessing the effects of additional environmental stresses on the yield and composition of peppermint essential oil. An analogous analysis of transgenic peppermint plants with modulated expression levels of monoterpene biosynthetic genes was also deemed highly beneficial for increasing the robustness and our modeling and its utility in guiding efforts aimed at modulating essential oil yield and composition. Mahmoud and Croteau [7] reported that, by over-expressing the gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), an up to 1.5-fold essential oil yield increase was observed. Antisense suppression of the (+)-menthofuran synthase (MFS) gene led to a dramatic decrease in the amounts of the undesirable side product (+)-menthofuran [7]. A slight increase in overall monoterpene yields was reported for transgenic plants with increased expression levels of the gene encoding (-)-limonene synthase (LS) [8], whereas only minor effects on yield were detected in an independent study [9]. Transgenic plants over-expressing the gene coding for (-)-limonene 3-hydroxylase (L3H) did not accumulate increased levels of the recombinant protein and the composition and yield of the essential oils was



**Fig. 4.1.** Outline of *p*-menthane monoterpene biosynthesis in peppermint glandular trichomes. The following enzymes are involved in this pathway: 1, 1-deoxy-D-xylulose 5-phosphate synthase; 2, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; 3, 2C-methyl-D-erythritol 4-phosphate cytidyltransferase; 4, 4-(cytidine 5'-diphospho)-2C-methyl-D-erythritol 4-phosphate kinase; 5, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; 6, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate synthase; 7, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase; 8, isopentenyl diphosphate isomerase; 9, geranyl diphosphate synthase; 10, (-)-limonene synthase; 11, (-)-limonene 3-hydroxylase; 12, (-)-trans-isopiperitenol dehydrogenase; 13, (-)-trans-isopiperitenone reductase; 14, (+)-cis-isopulegone isomerase; 15, (+)-menthofuran synthase; 16a, (+)-pulegone reductase ((-)-menthone-forming activity); 16b, (+)-pulegone reductase ((+)-isomenthone-forming activity); 17a, (-)-menthone (-)-menthol reductase ((-)-menthol-forming activity); 17b, (-)-menthone (-)-menthol reductase ((+)-neoisomenthol-forming activity); 18a, (-)-menthone (+)-neomenthol reductase ((+)-neomenthol-forming activity); 18b, (-)-menthone (+)-neomenthol reductase ((+)-isomenthol-forming activity). The subcellular compartmentation of *p*-menthane metabolic enzymes is color-coded: Cyt (blue), cytosol; ER (orange), endoplasmic reticulum; Lpl (green), leucoplasts; Mit (red), mitochondria. The inhibitory effects of (+)-menthofuran and geranyl diphosphate on (+)-pulegone reductase and isopentenyl diphosphate isomerase, respectively, is indicated by red arcs with orthogonal red lines. Names of selected metabolites are shown in the colors that are used to indicate the corresponding profiles in Fig. 4.2, 4.3 and 4.5.

the same as in wild-type controls; however, co-suppression of the L3H gene resulted in a vastly increased accumulation of the intermediate (-)-limonene, without notable effects on oil yield [10].

In the present study we have reexamined some of these transgenic lines in order to obtain essential data for understanding the factors controlling peppermint oil yield. During routine analyses of transgenic lines that had been propagated in the greenhouse for 7 years, we detected significantly elevated essential oil quantities, when compared to wild-type controls, in the MFS7 antisense line. This surprising finding prompted us to investigate the mechanisms determining oil yield. Our data indicate that genotype-dependent and environmental effects on essential oil yield correlate directly with the density and size distribution of glandular trichomes on the leaf surface.

## **4.2 Materials and Methods**

### **4.2.1 Chemicals, plant material and growth conditions**

All chemicals were purchased from Sigma-Aldrich. Peppermint (*Mentha x piperita* cv. Black Mitchum) plants were grown on soil (Sunshine Mix LC1, SunGro Horticulture) in a greenhouse with supplemental lighting from sodium vapor lights ( $850 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation at plant canopy level) with a 16 h photoperiod and a temperature cycle of 27°C/21°C (day/night). Transgenic plants were provided by the laboratory of Dr. R. Croteau (WSU). The initial characterization of these transgenic lines was published previously [7,10]. Plants were watered daily with a fertilizer mix (N:P:K

20:20:20, v/v/v; plus iron chelate and micronutrients). Stress experiments were performed by (1) reducing water amounts (50 % of the regular volume), (2) moving plants to a growth chamber with a 16 h photoperiod at reduced light levels ( $300 \mu\text{mol m}^{-2} \text{s}^{-1}$  of photosynthetically active radiation at plant canopy level), and (3) combining a low light treatment (as above) with high night temperatures (30°C/30°C; day/night).

#### **4.2.2 Monoterpene analysis**

Leaves were directly (without prior freezing) steam-distilled and solvent-extracted using 10 mL of pentane in a condenser-cooled Likens-Nickerson apparatus [11]. Monoterpenes were identified by comparison of retention times and mass spectra to those of authentic standards in gas chromatography with mass spectrometry detection. Quantification was achieved by gas chromatography with flame ionization detection based upon calibration curves with known amounts of authentic standards and normalization to the peak area of camphor as internal standard.

#### **4.2.3 Glandular trichome distribution**

The distribution of glandular trichomes on peppermint leaves was evaluated using the method described by Turner et al. [12] with minor modifications. Briefly, leaves were cut along their blade and each half was divided into three sampling zones (basal, middle and apical). Both abaxial and adaxial leaf surfaces were sampled. Transmission Electron Microscopy grids (50 mesh, 3 mm diameter; containing 12 grid squares with an enclosed

area of about 0.180625 mm<sup>2</sup> each; Pelco International) were placed on leaf surfaces. Glandular trichome counting was performed in five grids per zone and on five different leaves. The total leaf area and the diameters of individual glandular trichomes were calculated based on digitized images of leaves (ImageJ; open source software developed by the National Institutes of Health) using previously described methods [12]. The calculations of essential oil volume per trichome were performed as described in Rios-Esteba et al. [6].

#### **4.2.4 Real-time quantitative PCR**

Oil gland secretory cells were isolated from peppermint leaves at 15 and 30 d after leaf emergence using a previously published protocol [13]. Total RNA was extracted from isolated secretory cells using the Trizol® Reagent (Gibco BRL) according to the manufacturer's instructions. Total RNA (1 µg) was treated with RNase-free DNase (Fermentas Life Science) and first-strand cDNA was synthesized using reverse transcriptase (Invitrogen). RNA isolation and cDNA synthesis were carried out with three independent biological replicate samples. In a 10 µL qPCR reaction, concentrations were adjusted to 18 nM (primers), 2.5 µM (oligo dTs) and 0.5 mM (dNTPs). Reactions were performed in a 96-well optical plate at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60° for 10 min (7500 Real-Time PCR system (Applied Biosystems)). Fluorescence intensities of three independent measurements were normalized against the ROX reference dye. For each experimental sample, the amount of target and endogenous control (β-actin gene; AW255057) were determined using the comparative C<sub>T</sub> method

according to the manufacturer's instructions (Applied Biosystems). The calculated average target value was subtracted from the average endogenous control value to obtain  $\Delta C_T$ . First-strand cDNA obtained from RNA of wild-type secretory cells (harvested at late post-secretory stage and thus with very low expression levels of target genes) was used as the template for calibrator measurements as above. The  $\Delta C_T$  of the calibrator was then subtracted from the  $\Delta C_T$  of the target to obtain  $-\Delta\Delta C_T$ . The relative amount of target, normalized to an endogenous control and relative to the calibrator, was then calculated as  $RQ = 2^{-\Delta\Delta C_T}$ . Thus, the normalized amount of each target is expressed as a unitless number and all quantities are expressed as an x-fold difference relative to the value obtained with the calibrator sample. Primer design for the genes of interest was performed using the Primer Express software (Applied Biosystems). The following primer pairs (obtained from Sigma-Genosys) were used (gene numbering as in Fig. 4.1):

1, AF019383, Forward (F): 5'-AAGAAGCAGGGCTGAGTCCTAA-3'; Reverse (R): 5'-TGTCCTTTCCTCCTCCAATCA-3';

2, AF116825, F: 5'-CGGCTACCTCGACATTTTCAA-3'; R: 5'-GCGACACCGCCATTTCC-3';

4, AF179283, F: 5'-TGTCCCACTGGTGAAGTTTACAA-3'; R: 5'-CTAGCAACACCAATGGATCAATG-3';

6, AW255909, F: 5'-TCGACCTGTATGTTGGCA-3'; R: 5'-GATTAGGGCATCCGTAGCATTC-3';

10, AW255536, F: 5'-CTCGCACTCAACAACCTTCGTC-3'; R: 5'-CCCACGATTGTCGAAGATAGG-3';

11, AF124817, F: 5'-CGTTCGGAGCGGGAAGA-3'; R: 5'-CAATGGAACCTCAACGTTTGC-3';

15, EU108704, F: 5'-GCCGGAACCGATACGACTTT-3'; R: 5'-TTTTAGGGTACGCGGGTTTTT-3';

16, AY300163, F: 5'-



GGAAGATGCTTGAAGCTGTGATC-3'; R: 5'-ACTGGGAGACCATCCCACATAC-3'.

#### 4.2.5 Kinetic modeling

Descriptions of modeling assumptions, variables, parameters, and the MATLAB source code, are provided in MIRIAM-compatible format [14] in Supplementary Material. Kinetic mathematical modeling to simulate monoterpene profiles in peppermint leaves was performed according to [6], with several modifications. Briefly, we expanded the model to include the methylerythritol 4-phosphate (MEP) pathway, which provides the precursors for peppermint monoterpene essential oil biosynthesis [15]. Our modified model also does not contain the factors  $p$  and  $q$ , which we had introduced to account for the fraction of (+)-pulegone reductase (PR) activity affected by feedback ((+)-menthofuran) and substrate ((+)-pulegone) inhibition, respectively. Instead, a new factor  $w$  was introduced, which allows us to account for the very small amounts of the intermediate (+)-pulegone retained within the secretory cells (not secreted into the subcuticular oil storage cavity). We continue to utilize the factor  $z$ , which accounts for the selective retention of (+)-menthofuran in secretory cells under stress conditions. Furthermore, based on the gene expression data presented here, the enzyme concentrations used in our model were adjusted for different peppermint genotypes grown under various environmental conditions (for details see Supplementary Material). A function that accounts for the developmental changes in the density of glandular trichomes was also added to the code of our model. Thus, simulations of essential oil

yield and composition in individual glandular trichomes can now be extrapolated to the scale of an entire leaf (for details see Supplementary Material).

## **4.3 Results and Discussion**

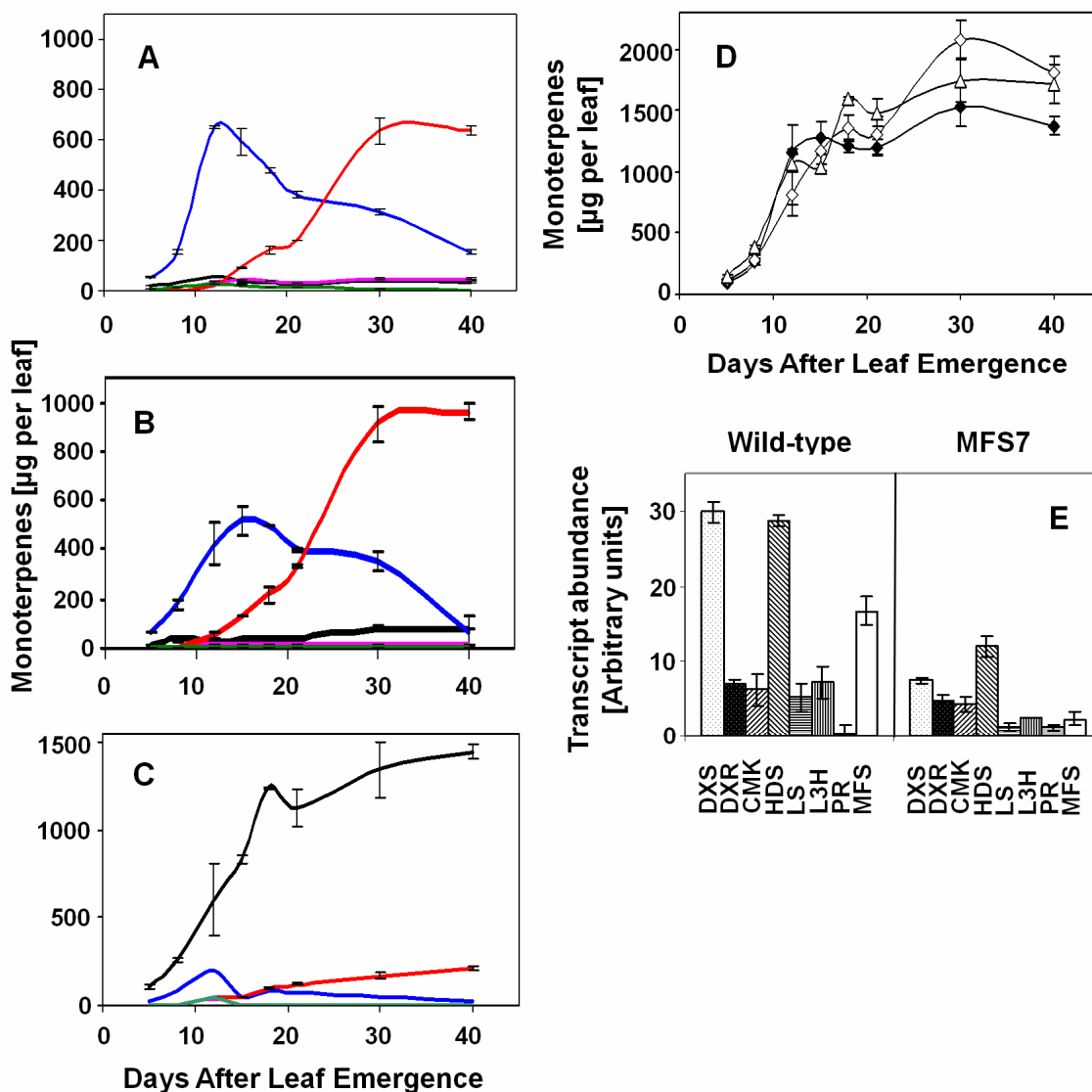
### **4.3.1 Transgenic MFS7 plants produce higher monoterpenoid essential oil yields than wild-type controls**

As part of our ongoing effort to expand our capabilities to utilize mathematical modeling as a tool for generating testable hypotheses regarding the regulation of metabolic pathways, we evaluated the plasticity of monoterpene composition in transgenic peppermint plants. In particular, we chose to reevaluate two transgenic lines that had previously been demonstrated to produce monoterpene profiles that are significantly different from those of wild-type plants. One of these lines, designated MFS7, which had been transformed with a construct containing an antisense cDNA of the peppermint (+)-menthofuran synthase gene [7], was generated initially to reduce the levels of the undesirable dead-end pathway side product (+)-menthofuran. The second transgenic line (designated L3H20 or pGALH20) contained, because of cosuppression effects of the expressed transgene, decreased levels of (-)-limonene 3-hydroxylase transcript [10]. These transgenic plant lines were propagated under greenhouse growth conditions for 7 years (MFS7) and 4 years (L3H20), respectively. To enable kinetic mathematical modeling, we obtained time course data of monoterpene accumulation from leaf emergence to maturity. The MFS7 antisense plants accumulated, compared to wild-

type plants, lower levels of both (+)-menthofuran (0.2-fold of WT) and (+)-pulegone (0.2-fold of WT) (Fig. 4.2A,B). These measurements were in good agreement with previously published results [7]. Surprisingly, we also measured higher levels of (-)-limonene (1.5-fold up at 40 d) and increased total monoterpenoid essential oil yields in MFS7 (1.35-fold up at 30 and 40 d) compared to wild-type plants (Fig. 4.2D). The L3H20 line accumulated vastly increased levels of (-)-limonene (approx. 80 % of all monoterpenes) (Fig. 4.2C), whereas the total monoterpene content in this line was similar to that of wild-type plants (Fig. 4.2D), which is consistent with previously published data [10].

#### **4.3.2 Biosynthetic gene expression patterns correlate with monoterpenoid essential oil composition but not with yield in transgenic MFS7 plants**

To assess the reasons for increased oil yields in the transgenic line MFS7, oil gland secretory cells, which are responsible for the synthesis of essential oils in peppermint, were isolated from leaves at 15 d after leaf emergence (the time of maximum essential oil biosynthetic activity), RNA was extracted (modified from [13]), and the expression levels of key genes involved in determining oil quantity and composition were assayed using qPCR. In peppermint, precursors for monoterpenoid essential oils are synthesized via the plastidial mevalonate-independent pathway [15]. The expression levels of the genes encoding 1-deoxy-D-xylulose 5-phosphate synthase (DXS), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), 4-diphosphocytidyl-2C-methyl-D-erythritol kinase (CMK), and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (HDS) were



**Fig. 4.2.** Experimentally determined monoterpene profiles of greenhouse-grown wild-type (A), MFS7 transgenic (B), and L3H20 transgenic (C) peppermint plants. The following colors are used for indicating monoterpene profiles: (-)-limonene, black; (+)-pulegone, pink; (+)-menthofuran, green; (-)-menthone, blue; (-)-menthol, red. Total monoterpene yield in wild-type (black diamonds), MFS7 transgenic (white diamonds), and L3H20 transgenic (white triangles) peppermint plants are shown in panel D. Panel E depicts expression patterns of genes involved in peppermint monoterpene biosynthesis, as determined by real-time quantitative PCR, using the peppermint  $\beta$ -actin gene (AW255057) as an endogenous control. The average signal intensity of RNA obtained with 30 d samples (wild-type plants grown under greenhouse conditions) was used as a calibrator (based on prior knowledge expression levels of genes involved in monoterpene biosynthesis are consistently low (but detectable) at this stage of leaf development). The following abbreviations and acronyms are used: DXS, -deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMK, 4, 4-(cytidine 5'-diphospho)-2C-methyl-D-erythritol 4-phosphate kinase; HDS, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate synthase; LS, (-)-limonene synthase; L3H, (-)-limonene 3-hydroxylase; PR, (+)-pulegone reductase; MFS, (+)-menthofuran synthase.

4.0-fold, 1.4-fold higher, 1.4-fold and 2.4-fold higher in wild-type controls compared to MFS7 plants, respectively (Fig. 4.2E). The increased oil yield in MFS7 plants (compared to wild-type) was thus not reflected in the expression patterns of genes that code for the enzymes involved in precursor supply to the monoterpene pathway

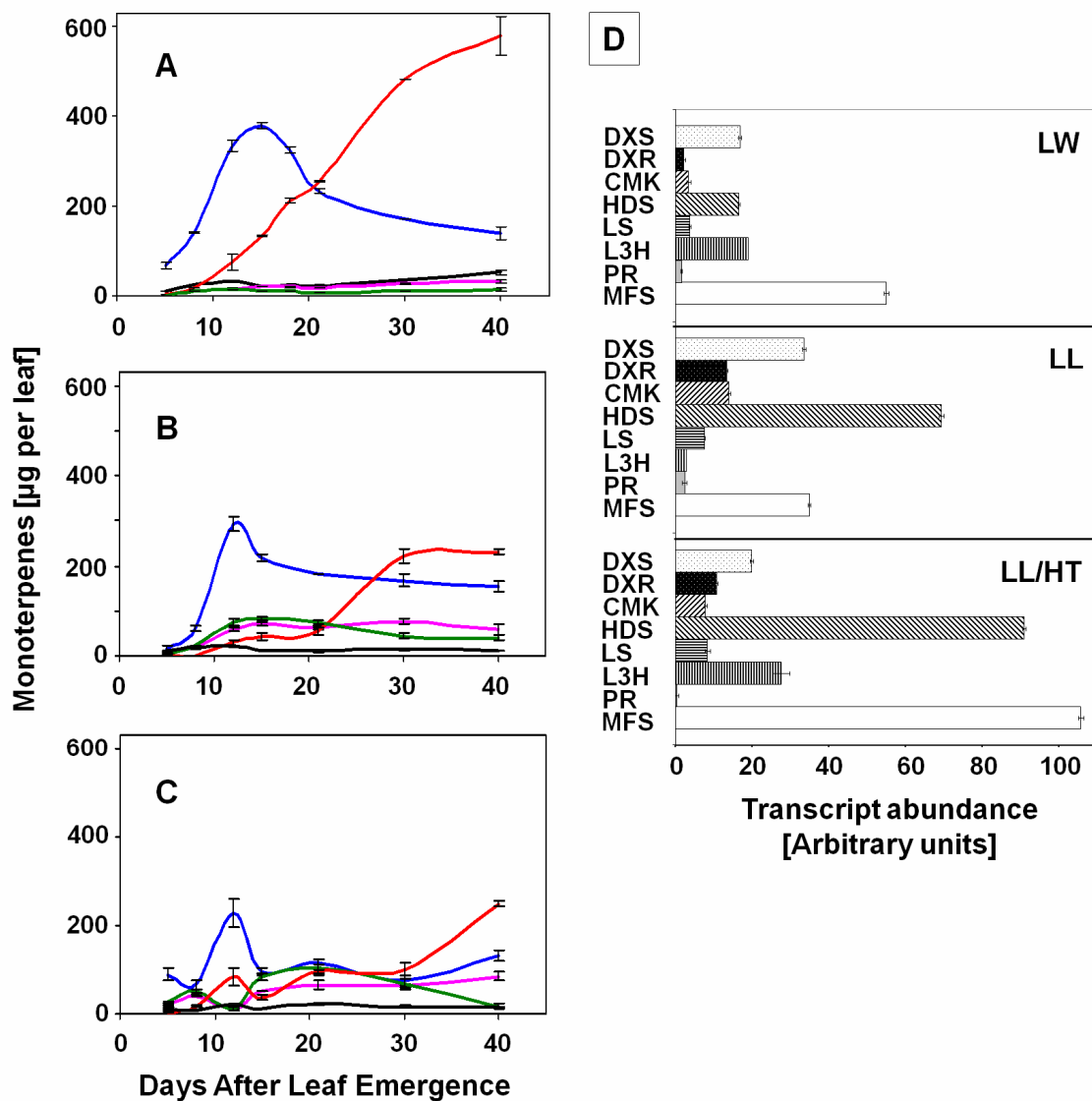
In wild-type plants, the composition of the essential oil is determined primarily by the monoterpene pathway-specific enzymes [5]. The genes encoding (-)-limonene synthase (LS), (-)-limonene 3-hydroxylase (L3H) and (+)-menthofuran synthase (MFS) were expressed at high levels in controls (4.4-fold, 2.9-fold, and 7.2-fold up compared to MFS7 plants), whereas the (+)-pulegone reductase (PR) gene was expressed at very low levels (6.9-fold down compared to MFS7 plants). These expression patterns did not provide an indication why increased essential oil yields were detected in MFS7 plants. In contrast, the decreased amounts of (+)-pulegone and (+)-menthofuran in MFS7 plants (compared to wild-type) were indeed reflected in the expression levels of the biosynthetic genes (low MFS and high PR expression levels).

### **4.3.3 Biosynthetic gene expression patterns correlate with monoterpenoid essential oil composition but not with yield under stress conditions**

The yield and composition of peppermint essential oil is greatly influenced by environmental factors. When plants are grown under low light intensities, water deficit or high night temperatures, poor quality oils with elevated amounts of the undesirable metabolites (+)-pulegone and (+)-menthofuran are produced and decreases in oil yields have been reported [16,17]. In a previous paper we demonstrated, using an iterative

approach of mathematical modeling and experimental testing, that compositional changes can be explained by a stress-induced increase of MFS transcript abundance and enzyme activity, and inhibitory effects of (+)-menthofuran on PR activity (competitive against (+)-pulegone) [6]. In the present paper we assessed the potential mechanisms for variation in peppermint monoterpenoid essential oil yields. Plants were grown in the greenhouse with supplemental lighting (controls), under water deficit (50 % of regular volume of fertilizer mix), at decreased light intensity (roughly 1/3 of regular photon flux), and under conditions with a combination of low light intensities (as above) and high night temperatures (constant day and night temperatures at 30°C) (for details see Materials and Methods).

Under drought conditions, the profiles of the major monoterpenes changed only slightly (Fig. 4.3A) but the overall monoterpenoid essential oil yield decreased 1.6-fold compared to plants grown in the greenhouse (Fig. 4.3D). In low-light-grown plants, the amounts of the intermediate (+)-pulegone increased transiently to 82 µg/leaf (4.1-fold higher than greenhouse-grown controls), whereas the (+)-menthofuran content increased linearly to 77 µg/leaf at 30 d (1.7-fold higher than greenhouse-grown controls) (Fig. 4.3B). In contrast, MFS7 plants accumulated (+)-pulegone and (+)-menthofuran only to low levels (2 and 23 µg/leaf, respectively), even when grown under low light conditions. The total monoterpene yield of low-light grown wild-type plants (658 µg/leaf) was 2.3-fold lower than in plants cultivated in the greenhouse (Table 4.1). In MFS7 plants the total monoterpene yield under low light conditions was 1001 µg/leaf (1.5-fold higher than in wild-type plants), corresponding to a 2.1-fold decreased compared to greenhouse-grown MFS7 plants.



**Fig. 4.3.** Experimentally determined monoterpene profiles of wild-type plants grown under various stress conditions: (A), low water; (B), low light intensity; (C), low light intensity and high night temperatures. Panel D depicts expression patterns of genes involved in peppermint monoterpene biosynthesis, as determined by real-time quantitative PCR, using the peppermint  $\beta$ -actin gene (AW255057) as an endogenous control. For legend see Fig. 4.2.

**Table 4.1.** Glandular trichome density and size distribution at 30 d after leaf emergence.  
For details see text.

<i>Experimental Condition</i>	<i>Glandular Trichomes</i>			<i>Monoterpene Yield</i>		
	<i>Size</i>	<i>Distribution [%]</i>	<i>Density [per leaf]</i>	<i>Calculated [ηl per leaf]</i>	<i>Calculated [ηg per leaf]</i>	<i>Experimental [ηg per leaf]</i>
Wild-type (Greenhouse)	<i>Large</i>	39	3984	0.533995	481	
	<i>Average</i>	57	5823	0.514422	463	
	<i>Small</i>	4	409	0.023102	21	
	<i>Total</i>		10215	1.071520	964	1535 ± 114
Wild-type (Low water)	<i>Large</i>	44	3200	0.428944	386	
	<i>Average</i>	53	3855	0.340562	307	
	<i>Small</i>	3	218	0.012336	11	
	<i>Total</i>		7273	0.781842	704	974 ± 51
Wild-type (Low light)	<i>Large</i>	40	2802	0.375526	338	
	<i>Average</i>	56	3922	0.346530	312	
	<i>Small</i>	4	280	0.015840	14	
	<i>Total</i>		7004	0.737897	664	658 ± 73
Wild-type (Low light and high night temp.)	<i>Large</i>	7	369	0.049485	45	
	<i>Average</i>	49	2584	0.228319	205	
	<i>Small</i>	44	2321	0.131204	118	
	<i>Total</i>		5274	0.409007	368	377 ± 9
MFS7 Transgenics (Greenhouse)	<i>Large</i>	67	369	1.030443	927	
	<i>Average</i>	33	2584	0.334532	301	
	<i>Small</i>	0	2321	0	0	
	<i>Total</i>		11474	1.364975	1228	2079 ± 111



When low light stress was combined with high night temperatures, the total monoterpenoid essential oil yield in wild-type plants dropped dramatically to 377  $\mu\text{g}/\text{leaf}$  at 30 d (4.1-fold decrease compared to greenhouse-grown controls) (Table 4.1). Under the same growth conditions, the monoterpene yield in MFS7 plants was 445  $\mu\text{g}/\text{leaf}$  (1.2-fold higher than in wild-type plants). In wild-type plants grown under low light and high night temperature conditions, the amounts of the intermediate (+)-pulegone increased transiently to 105  $\mu\text{g}/\text{leaf}$  (5.3-fold higher than greenhouse-grown controls) and the (+)-menthofuran content increased linearly to 63  $\mu\text{g}/\text{leaf}$  at 30 d (1.4-fold higher than greenhouse-grown controls) (Fig. 4.3C). For both (+)-pulegone and (+)-menthofuran the measured levels in MFS7 plants grown under the same conditions were much lower (13 and 63  $\mu\text{g}/\text{leaf}$ , respectively). In general, these results confirmed those obtained in a previous study with wild-type and MFS7 plants [7]: even under stress conditions, the antisense suppression of MFS leads to low accumulation levels of the undesirable intermediates (+)-pulegone and (+)-menthofuran. However, in the present study, we also detected higher monoterpenoid essential oil yields in MFS7 plants (when compared to wild-type controls) under various environmental conditions.

To assess the role of transcriptional regulation in determining monoterpene profiles we quantified the expression levels of key biosynthetic genes using qPCR. The transcript level patterns of DXS, DXR, CMK, HDS, and LS were very similar. In comparison to greenhouse-grown (GH) controls their expression levels decreased under drought (LW) and increased under low light (LL) and when grown at low light/high night temperatures (LL/HT) (Fig. 4.3D). The expression patterns of the L3H gene were lower in LL (2.5-fold down) and higher in LW and LL/HT samples (2.6- and 3.9-fold up, respectively)

compared to GH controls (Fig. 4.3D). The expression levels of the genes encoding PR and MFS increased under all stress conditions. In general, the various stress treatments resulted in similar – but less pronounced – responses, in MFS7 plants (data not shown). Major differences were detected in the expression levels of the genes encoding PR (lower in wild-type vs. MFS7 under all conditions) and MFS (higher in wild-type vs. MFS7 under all conditions), which explains the measured differences in monoterpene composition (higher (+)-pulegone and (+)-menthofuran in wild-type vs. MFS7 plants). However, the differences in total monoterpenoid essential oil yield between wild-type and MFS7 plants and both genotypes under different environmental conditions, are not reflected in monoterpene biosynthetic gene expression patterns.

#### **4.3.4 Monoterpenoid essential oil yields correlate with glandular trichome density and size distribution**

Besides differences in the expression levels of genes involved in monoterpene biosynthesis, differences in the storage capacity or density of glandular trichomes, the anatomical structures harboring essential oils, could also account for variations in essential oil yields. Trichomes were divided into three different classes: large ( $> 75 \mu\text{m}$  diameter), average (65-74  $\mu\text{m}$  diameter) and small ( $< 65 \mu\text{m}$  diameter). At 30 d after leaf emergence the majority of glandular trichomes on WT leaves was of average size, with a substantial proportion (39 %) of large-sized and a low proportion (4 %) of small-sized trichomes (Table 4.1). In contrast, the MFS7 line contained a substantially higher proportion of large-sized glandular trichomes (67 % at 30 d) (Table 4.1). The average

WT leaf at 30 d contained 10,215 glandular trichomes. Based on size each trichome was estimated to contain  $1.34 \cdot 10^{-4}$   $\mu\text{l}$  (large-sized trichomes; calculated by approximating the shape of the oil storage cavity as a hemisphere; for details see Supplementary Material),  $8.84 \cdot 10^{-5}$   $\mu\text{l}$  (average-sized trichomes) or  $5.65 \cdot 10^{-4}$   $\mu\text{l}$  (small-sized trichomes). When glandular trichome numbers, distribution and volume were taken into account, the total monoterpene yield at 30 d was calculated to be 964  $\mu\text{g}$  per leaf for WT plants, which was roughly 1.6-fold lower than the experimentally determined value of  $1,535 \pm 114$   $\mu\text{g}$  per leaf (Table 4.1). At 30 d MFS7 plants contained 11,474 glandular trichomes per leaf. Using the same approach as described above, total monoterpene content was estimated to be 1,228  $\mu\text{g}$  per leaf (1.6-fold lower than the experimental measurement of  $2,079 \pm 111$   $\mu\text{g}$  per leaf) (Table 4.1). The calculated monoterpene contents, based on glandular trichome number and size distribution, were about 1.32-fold higher in MFS7 than in WT, which was very close to the measured 1.35-fold difference in monoterpene yields.

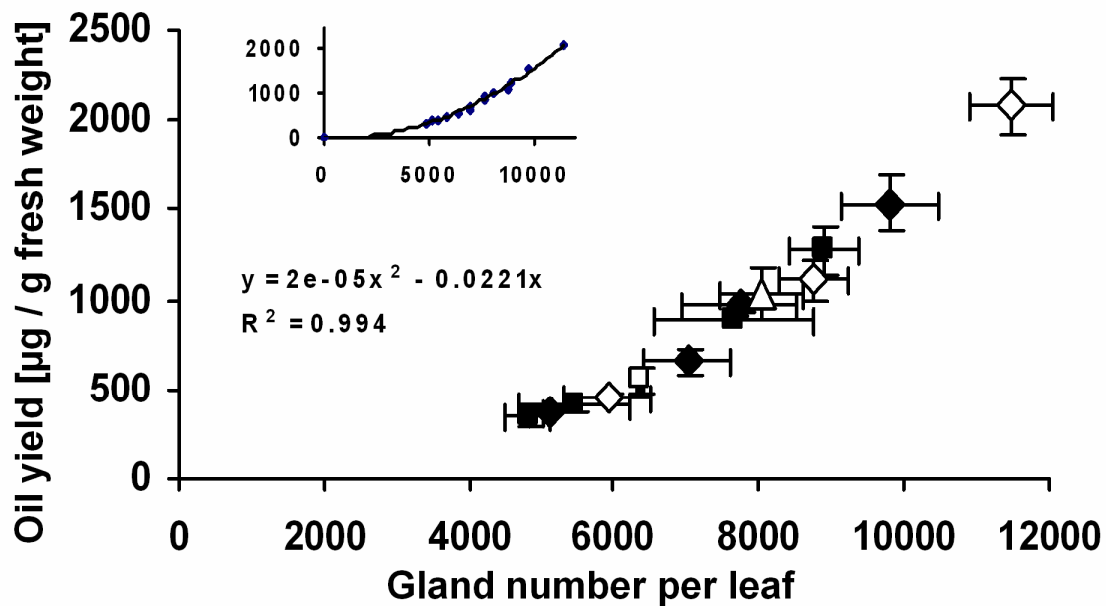
Plants grown under water deficit conditions produced  $974 \pm 11$   $\mu\text{g}$  total monoterpenes per leaf, corresponding to a 1.6-fold decrease compared to greenhouse-grown controls. A concomitant decrease was also observed for the number of glandular trichomes (7,273 per leaf) and the calculated essential oil yield based on glandular trichome counts and volume estimations (Table 4.1). When peppermint plants were grown under low light intensities, the essential oil yield ( $658 \pm 47$   $\mu\text{g}$  monoterpenes per leaf) was roughly 2.3-fold lower than in controls and could be accounted for by lower glandular trichome numbers (Table 4.1). Under severe stress (combination of low light intensity and high night temperatures), the essential oil yield was even lower ( $377 \pm 55$   $\mu\text{g}$  monoterpenes

per leaf). The number of glandular trichomes per leaf in stressed plants was lower than in controls and the size distribution was shifted toward a higher percentage of smaller glandular trichomes (Table 4.1).

Interestingly, when leaf trichome density, as determined with different genotypes (wild-type as well as transgenic plants MFS7 and L3H20) and under different growth conditions, was plotted against monoterpenoid essential oil yields, the resulting line graph could be fitted to the shape of a quadratic (second order) polynomial function ( $R^2 = 0.994$ ; Fig. 4.4). Taken together, these results indicate that the primary determinant of monoterpenoid essential oil yield in peppermint is the density of glandular trichomes. The size distribution of glandular trichomes, although important, remains a secondary determinant of oil yield.

#### **4.3.5 Mathematical modeling of leaf monoterpenoid essential oil profiles**

At face value the gene expression patterns we obtained with peppermint wild-type and transgenic plants grown under various environmental conditions appeared to support a key role for transcriptional control in determining monoterpenoid essential oil composition. As demonstrated in an earlier study, post-translational regulation (competitive inhibition of PR by the dead-end side product (+)-menthofuran) is also an important mechanism controlling oil composition under stress conditions [6]. To evaluate if a combination of gene expression patterns and glandular trichome density and size distribution could account for both monoterpene composition and yield in wild-type and MFS7 plants under different environmental conditions, we modified our existing



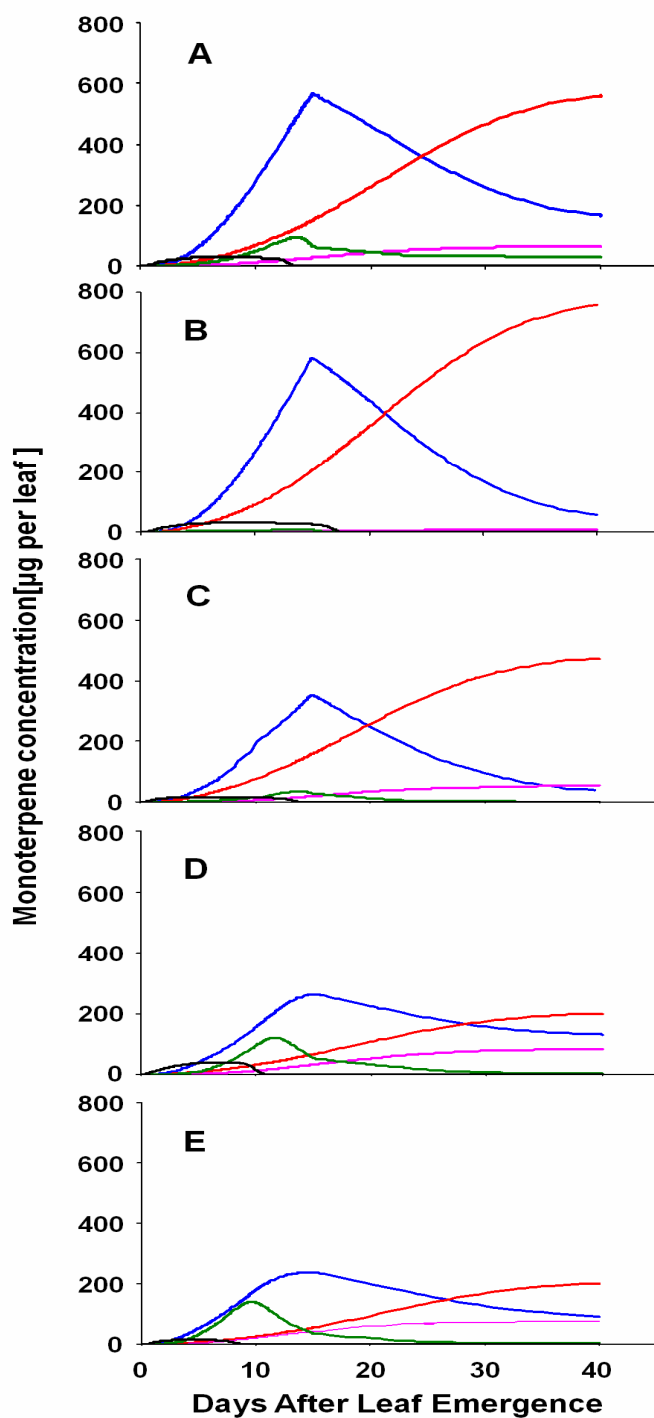
**Fig 4.4.** Correlation of glandular trichome density and monoterpenoid essential oil yield in peppermint leaves. Symbols: black rectangles, wild-type, 15-d-old- leaves, different environmental conditions; black diamonds, wild-type, 30-d-old- leaves, different environmental conditions; white rectangles, MFS7 line, 15-d-old- leaves, different environmental conditions; white diamonds, MFS7 line, 30-d-old- leaves, different environmental conditions; white triangles, L3H20 line, 30-d-old- leaves grown under greenhouse conditions. The inset shows averaged data points with a curve fitted using a quadratic (second order) polynomial function.

kinetic mathematical model for peppermint essential oil biosynthesis with the data obtained as part of the present study: (1) since essential oil yield might be affected by the enzymatic steps that provide precursors for monoterpene biosynthesis, we expanded our model to include the plastidial MEP pathway, which is the sole carbon source for monoterpene biosynthesis in peppermint [15]; for details see Table 1 of Supplementary Material); (2) initial enzyme concentrations were adjusted based on gene expression patterns determined experimentally with different peppermint genotypes under various environmental conditions (for details see Figs. 4.2E and 4.3D of manuscript and Table 2 of Supplementary Material); (3) a logistic function (sigmoid curve) was added to account for the developmental dynamics of glandular trichome density (for details see Supplementary Material).

Our initial simulations of monoterpene profiles for plants grown under different environmental conditions were performed by varying enzyme concentrations, based on experimental data on gene expression patterns, in rate expressions for Michaelis-Menten kinetics. We used a filter based on the assumption that enzyme concentrations vary much less than gene expression levels. For example, a 5-fold upregulation at the steady-state transcript level was converted into a 1.3 to 1.6-fold increase in enzyme concentration (Table 2 in Supplementary Material). The simulated monoterpene compositions at 40 d after leaf emergence were quite similar to those obtained experimentally by steam-distillation. However, the shapes of the monoterpene curves and the essential oil yields were very different from experimental values (data not shown). Based on our observation that glandular trichome density correlated strongly with essential oil yield (Fig. 4.4), we tested several functions to approximate the developmental changes in

trichome distribution. The best results were obtained with a logistic function (for details see Supplementary Material) (Fig. 4.5). This function specifies an initial lag time, then grows quickly but, because of limits in the size of the glandular trichome population, eventually levels off at the later stages of leaf development. Without any further assumptions, the shapes of simulated monoterpene curves matched experimentally obtained profiles very closely. In particular, high yields of (-)-menthol, with very low levels of (+)-pulegone and (+)-menthofuran, were simulated for MFS7 plants (Fig. 4.5B). The essential oil yield simulated for plants grown under drought conditions was lower than that of fully watered plants but the general profile did not change significantly (Fig. 4.5C). Under low light and low light/high night temperature conditions very low essential oil yield and high amounts of (+)-menthofuran were simulated (Fig. 4.5D, E).

Taken together, these simulations indicate that we can account for all major factors affecting essential oil yield and composition in wild-type plants and selected transgenic lines. Mathematical modeling can thus be employed to guide metabolic engineering efforts aimed at modulating monoterpene essential oil profiles in peppermint leaves. We are now in the position to test, based on a combination of mathematical modeling and experimental testing, the utility of peppermint glandular trichomes as cellular factories for the production of high value terpenoids.



**Fig 4.5.** Computer simulation of monoterpene profiles for plants grown under different environmental conditions. Modeling assumptions are based on experimental results presented in this study (for details see Suppl. Material). Color-coding of monoterpene profiles as in Fig. 4.2.



## Acknowledgements

The authors would like to thank Dr. Rodney Croteau (Washington State University) for permission to use transgenic plants (lines MFS7 and L3H20), which had been generated in his laboratory. We would also like to thank Julia Gothard-Szamosfalvi and Craig Whitney for growing plants and assistance with operating growth chambers. R.R.E. thanks the Fulbright program and the University of Antioquia (Medellin, Colombia) for scholarships. This work was supported in part by the Agricultural Research Center at Washington State University and the Mint Growers of the Northwest United States.

## 4.4 References

- [1] J. Rohloff, *Agric. Food Chem.* 47 (1999) 3782-3786.
- [2] J. Gershenzon, M. Maffei, R. Croteau, *Plant Physiol.* 89 (1989) 1351-1357.
- [3] D. McCaskill, J. Gershenzon, R. Croteau, *Planta* 187 (1992) 445-454.
- [4] F. Amelunxen, *Planta Med.* 13 (1965) 457-473.
- [5] R. Croteau, E.M. Davis, K.L. Ringer, M.R. Wildung, *Naturwiss* 92 (2005) 562-577.
- [6] R. Rios-Estepa, G.W. Turner, J.M. Lee, R.B. Croteau, B.M. Lange, *Proc. Natl. Acad. Sci. USA* 105 (2008) 2818-2823.
- [7] S.S. Mahmoud, R.B. Croteau, *Proc. Natl. Acad. Sci. USA* 98 (2001) 8915-8920.
- [8] F. Diemer, J.C. Caissard, S. Moja, J.C. Calchat, F. Jullien, *Plant Physiol. Biochem.* 39 (2001) 603-614.
- [9] S. Krasnyansky, R.A. May, A. Loskutov, T.M. Ball, K.C. Sink, *Theor. Appl. Genet.* 99 (1999) 676-682.
- [10] S.S. Mahmoud, M. Williams, R. Croteau, *Phytochemistry* 65 (2004) 547-554.

- [11] K.L. Ringer, M.E. McConkey, E.M. Davis, G.W. Rushing, R. Croteau, Arch. Biochem. Biophys. 418 (2003) 80-92.
- [12] G.W. Turner, J. Gershenzon, R. Croteau, Plant Physiol. 124 (2000) 655-664.
- [13] B.M. Lange, M.R. Wildung, E.J. Stauber, C. Sanchez, D. Pouchnik, R. Croteau, Proc. Natl. Acad. Sci. USA 97 (2000) 2934-2939.
- [14] N. Le Novère, A. Finney, M. Hucka, U.S. Bhalla, F. Campagne, J. Collado-Vides, E.J. Crampin, M. Halstead, E. Klipp, P. Mendes et al., Nat. Biotechnol. 23 (2005) 1509-1515.
- [15] W. Eisenreich, S. Sagner, M.H. Zenk, A. Bacher Tetrahedron Lett. 38 (1997) 3889-3892.
- [16] A.J. Burbott, W.D. Loomis Plant Physiol. 42 (1967) 20-28.
- [17] R.J. Clark, R.C. Menary Aust. J. Plant Physiol. 7 (1980) 685-692.

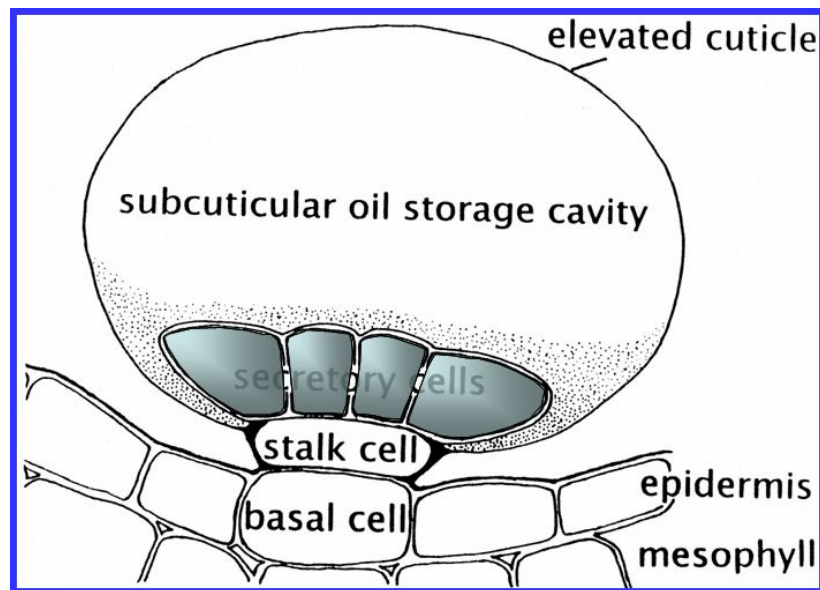
## 4.5 Supplementary Material

**Manuscript Title:** Monoterpenoid essential oil yield in peppermint is determined by the density and size distribution of glandular trichomes

### Modeling assumptions

Peppermint glandular trichomes harbor three different cells types (outlined in Scheme 1). Among these cell types the secretory cells (eight-celled disk highlighted in Scheme 1; only four cells are visible in cross section) are responsible for the biosynthesis of monoterpenes in peppermint.

### Scheme 1



- The shape of the secretory cell cluster was approximated by a frustum of a cone, the volume of which was calculated as  $\frac{1}{3} \pi h (R^2 + R r + r^2)$ .
- The volume of the essential oil-filled subcuticular cavity of mature glandular trichomes was calculated based on the approximation of its shape as a hemisphere

$(\frac{2}{3} \pi r^3)$ . Trichomes were divided into three different classes: large ( $> 75 \mu\text{m}$  diameter), average (65-74  $\mu\text{m}$  diameter) and small ( $< 65 \mu\text{m}$  diameter). The volumes were thus calculated as  $1.34 \cdot 10^{-4} \mu\text{l}$  (large-sized trichomes),  $8.84 \cdot 10^{-5} \mu\text{l}$  (average-sized trichomes) or  $5.65 \cdot 10^{-4} \mu\text{l}$  (small-sized trichomes).

- The volume densities of subcellular compartments were estimated as outlined in Rios-Esteba et al. (2008).
- Experimentally determined developmental patterns of biosynthetic enzyme activities were approximated using a Gaussian function (for more explanations see Rios-Esteba et al., 2008):

$$f(t) = \text{Comp} * \text{Enz} * \exp(-\frac{(t-b)^2}{2*c^2})$$

where:

Comp = Factor to adjust for the volume density of the compartment in which a particular enzyme is active

Enz = Enzyme concentration in the compartment in which it resides [uM]

t = Time [s]

b = Factor defining the position of the center of the Gaussian peak for enzyme activity [s]

c = Factor defining the width of the Gaussian peak for enzyme activity at half maximum [s]

- The following kinetic parameters were used in the mathematical model for greenhouse-grown wild-type plants:

**Table 1.** Kinetic parameters used in mathematical models for simulating monoterpene profiles of peppermint leaves.

Enzyme	Km [mM]	Kcat [s <sup>-1</sup> ]	Reference
(1) 1-Deoxy-D-xylulose 5-phosphate synthase (GAP)	0.068	1.9	Eubanks and Poulter (2003) Biochemistry 42, 1140-1149
(2) 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (Pyruvate)	0.44	1.9	Eubanks and Poulter (2003)
(2) 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXP)	0.132	4.4	Rohdich et al. (2006) FEBS J. 273, 4446-4458
(2) 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (MEP)	0.972	1.6	Rohdich et al. (2006) FEBS J. 273, 4446-4458
(3) 2C-Methyl-D-erythritol 4-phosphate cytidyltransferase	0.5	26	Rohdich et al. (2000) PNAS 97, 6451-6456
(4) 4-(Cytidine 5'-diphospho)-2C-methyl-D-erythritol 4-phosphate kinase	0.1*	1*	N.A.
(5) 2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase	0.252	1*	Rohdich et al. (2001) Eur. J. Biochem. 268, 3190-3197
(6) (E)-4-Hydroxy-3-methyl-but-2-enyl diphosphate synthase	0.42	0.4	Kollas et al. (2002) FEBS Lett. 532, 432
(7) (E)-4-Hydroxy-3-methyl-but-2-enyl diphosphate reductase	0.03	3.7	Altincicek et al. (2002) FEBS Lett. 532, 437 Graewert et al. (2004) J.Am.Chem.Soc. 126, 12847
(8) Isopentenyl diphosphate isomerase (DMAPP)	0.0051	0.018	Ramos-Valdivia et al. (1997) Eur.J.Biochem. 249, 161-170
(8) Isopentenyl diphosphate isomerase (IPP)	0.017	0.89	Ramos-Valdivia et al. (1997) Eur.J.Biochem. 249, 161-170
(9) Geranyl diphosphate synthase (DMAPP)	0.054	48	Burke et al. (1999) PNAS 96, 13062-13067
(9) Geranyl diphosphate synthase (IPP)	0.026	48	Burke et al. (1999) PNAS 96, 13062-13067
(10) (-)-Limonene synthase	0.020	0.3	Alonso et al. (1992) J. Biol. Chem. 267, 7582-7587
(11) (-)-Limonene 3-hydroxylase	0.018	1*	Karp et al. (1990) Arch. Biochem. Biophys. 276, 219-226
(12) (-)- <i>trans</i> -Isopiperitenol dehydrogenase	0.072	0.002	Ringer et al. (2005) Plant Physiol. 137, 863
(13) (-)-Isopiperitenone reductase	0.001	1.3	Ringer et al. (2003) Arch. Biochem. Biophys. 4186, 80-92
(14) (+)- <i>cis</i> -Isopulegone isomerase	0.27	1*	Kjonaas et al. (1985) Arch. Biochem. Biophys. 238, 49-60
(15) (+)-Menthofuran synthase	0.01*	0.9*	N.A.
(16) (+)-Pulegone reductase	0.0023	1.8	Ringer et al. (2003) Arch. Biochem. Biophys. 4186, 80-92
(17) (-)-Menthone:(-)-menthol reductase (menthone)	0.003	0.6	Davis et al. (2005) Plant Physiol. 137, 873-881
(17) (-)-Menthone:(-)-menthol reductase ((+)-isomenthone)	0.041	0.6*	Davis et al. (2005) Plant Physiol. 137, 873
(18) (-)-Menthone:(+)-neomenthol reductase (menthone)	0.674	0.06	Davis et al. (2005) Plant Physiol. 137, 873
(18) (-)-Menthone:(+)-neomenthol reductase ((+)-isomenthone)	1.0	0.06*	Davis et al. (2005) Plant Physiol. 137, 873

\* These values could not be obtained from the literature and have thus been estimated. All other kinetic values were reported in the papers cited above; N.A., not applicable.

Glandular trichome density and distribution depend on the developmental status of the leaf under investigation (see Table 4.1 of the manuscript). To account for developmental dynamics in trichome density, we introduced a logistic function, the most common

sigmoid curve. This function specifies an initial lag phase, then grows steeply but, because of limits in the size of the glandular trichome population, eventually levels off at the later stages of leaf development.

The expression for the logistic function is:

$$GN(t) = \frac{a}{1 + c \cdot e^{k \cdot t}}$$

Where

***a*** represents the limiting value of the output (in this case the number of glandular trichomes on a fully expanded leaf);

***c*** establishes a relationship between the initial and limiting output values (it represents the number of times the initial gland population must grow to reach ***a***);

***k*** is a factor determining the slope of the growth phase and time frame of the curve.

The code for the model that simulates monoterpene profiles of plants grown under greenhouse conditions is shown in full on pages 149-153. Table 2 on page 148 summarizes the changes to this reference model when monoterpene profiles are simulated for transgenic lines and plants grown under different environmental conditions. Experimentally determined differences of gene expression patterns (Table 2) and glandular trichome size and distribution (Fig. 4.4 of manuscript) are used to guide model modifications.

**Table 2. Model adjustments based on experimentally determined gene expression patterns.**

Enzyme	WT - GH <sup>a</sup> z=100 w=0.01 Gene Exp. Strd. [E]* in the model Level Error UM	MFS7 - GH z=10 w=0.01 Gene Exp. Strd. Fold Change [E]* in the model Level Error vs. WT - GH UM	WT - LL z=1500 w=0.01 Fold Change [E]* in the model vs. WT - GH UM	WT - LW z=400 w=0.01 Fold Change [E]* in the model vs. WT - GH UM	WT - LL/HT z=1500 w=0.01 Fold Change [E]* in the model vs. WT - GH UM
DXPS #	29.91 1.35 0.03	7.44 0.49 (-) 4.0 0.028	(+) 1.1 0.025	(-) 1.8 0.03	(-) 1.5 0.025
DXR	7.04 0.52 0.0225	4.87 0.55 (-) 1.5 0.0225	(+) 1.9 0.0225	(-) 3.2 0.0225	(+) 1.5 0.0225
CMK	6.19 2.21 0.0225	4.33 1.01 (-) 2.8 0.0225	(+) 2.3 0.0225	(-) 1.8 0.0225	(+) 1.3 0.0225
HDS	28.62 1.35 0.5	11.97 0.76 (-) 2.40 0.47	(+) 2.4 0.7	(-) 1.7 0.5	(+) 3.2 0.7
LS	5.18 1.89 0.017	1.19 0.53 (-) 4.3 0.015	(+) 1.5 0.017	(-) 1.4 0.017	(+) 1.7 0.017
L3H	7.18 2.18 0.003	2.49 0.02 (-) 2.9 0.0027	(-) 2.5 0.0027	(+) 2.6 0.004	(+) 3.9 0.004
PR	0.17 0.82 0.0015	1.14 0.66 (+) 6.9 0.002	(+) 14.3 0.0025	(+) 9.5 0.0025	(+) 1.7 0.0015
MFS	16.66 1.85 0.00007	2.31 0.87 (-) 7.2 0.00003	(+) 2.1 0.00009	(+) 3.3 0.00009	(+) 6.3 0.00009

\* Refers to the gaussian  $a$  parameter used to quantify the dynamic change of enzyme amount over time

# Gene/enzyme identifiers as in Fig. 1 of manuscript

<sup>a</sup> Abbreviations and acronyms: WT, wild-type; GH, greenhouse-grown; LL, grown under low light conditions in growth chamber; LW, grown under drought conditions in greenhouse; LL/HT, grown under low light and high night temperature conditions in growth chambers; MFS7, transgenic lines with reduced (+)-menthofuran synthase gene expression levels.

## Mathematical Model of Peppermint Monoterpenoid

### Essential Oil Biosynthesis

#### Reference model for wild-type plants grown under greenhouse conditions

```
function xdot = mint_MEP6_GH_WT(t,x)

% This function calculates monoterpene amounts (40 day time course) in leaves
% of peppermint WT plants grown in a greenhouse with supplemental lighting from sodium
% vapor lights.
% A mechanism following regular Michaelis-Menten-type kinetics is assumed for all
% enzymes with the following exceptions:
% (1) Substrate inhibition of (+)-pulegone reductase
% (2) Competitive inhibition of (+)-pulegone reductase by (+)-menthofuran
% (3) Competitive inhibition of isopentenyl-diphosphate isomerase by GPP
% (4) Reversible reaction mechanisms were assume for 1-Deoxy-D-xylulose-5-phosphate
% reductoisomerase and (-)-Menthone:(+)-neomenthol reductase
% (5) Bi-bi (two substrates, two products) reaction mechanisms were assumed for
% 1-deoxy-D-xylulose-5-phosphate synthase (Pyruvate + GAP = DXP + Co2) and geranyl
% diphosphate synthase (IPP + DMAPP = GPP + PPI). The former utilizes an ordered
mechanism
% (Pyr binds first), whereas a random mechanism is assumed for the latter.

% Metabolite Nomenclature

%[GAP]=x(1)          D-Glyceraldehyde 3-Phosphate
%[Pyr]=x(2)          Pyruvate
%[DOXP]=x(3)         1-Deoxy-D-xylulose 5-phosphate
%[ME4P]=x(4)         2-C-Methyl-D-erythritol-4-phosphate
%[CDPME]=x(5)        4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol
%[CDPME2P]=x(6)      2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
%[MEcPP]=x(7)        2-C-Methyl-D-erythritol-2,4-cyclodiphosphate
%[HMBPP]=x(8)        1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
%[DMAPP]=x(9)        Dimethylallyl-pyrophosphate
%[IPP]=x(10)         Isopentenyl diphosphate
%[GPP]=x(11)         Geranyl diphosphate
%[LIM]=x(12)         (-)-Limonene
%[IPPol]=x(13)       (-)-trans-Isopiperitenol
%[IPPone]=x(14)      (-)-Isopiperitenone
%[CIPUL]=x(15)       (+)-cis-Isopulegone
%[PUL]=x(16)         (+)-Pulegone
%[MF]=x(17)          (+)-Menthofuran
%[IMone]=x(18)       (+)-Isomenthone
%[Mone]=x(19)        (-)-Menthone
%[NMol]=x(20)        (+)-Neomenthol
%[Mol]=x(21)         (-)-Menthol
%[IMol]=x(22)        (+)-Isomenthol
%[NIMol]=x(23)       (+)-Neoisomenthol

% Kinetic Parameters

% kc units: [1/s]    kc = Kcat
% KM units: [uM]
% Ki units: [uM]

KM1a = 68;          %1-Deoxy-D-xylulose-5-phosphate synthase (DXS) for GAP
kc1a = 1.9;
KM1b = 440;        %1-Deoxy-D-xylulose-5-phosphate synthase (DXS) for Pyr
kc1b = 1.9;
Kia = 16;          %Dissociation constant for Pyr
KM2f = 132;        %1-Deoxy-D-xylulose-5-phosphate reductoisomerase (DXR; forward reaction)
kc2f = 4.4;
KM2r = 972;        %1-Deoxy-D-xylulose-5-phosphate reductoisomerase (DXR; reverse reaction)
kc2r = 1.6;
KM3 = 500;         %2-C-Methyl-D-erythritol 4-phosphate cytidyltransferase (MCT)
```



```

kc3 = 26;
KM4 = 100; %4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK)
kc4 = 1;
KM5 = 252; %2-C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPS)
kc5 = 1;
KM6 = 420; %4-Hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS)
kc6 = 0.4;
KM7 = 30; %4-Hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR)
kc7 = 3.7;
KM8f = 5.1; %Isopentenyl-diphosphate delta-isomerase for IPP (IPPI; forward reaction)
kc8f = 0.018;
KM8r = 17; %Isopentenyl-diphosphate delta-isomerase for DMAPP (IPPI; rev reaction)
kc8r = 0.89;
KM9a = 54; %Geranyl diphosphate synthase (GPPS; DMAPP as substrate)
kc9a = 48;
KM9b = 26; %Geranyl diphosphate synthase (GPPS; IPP as substrate)
kc9b = 48;
KM10 = 20; %(-)-Limonene synthase (LS)
kc10 = 0.3;
KM11 = 18; %(-)-Limonene 3-hydroxylase (L3H)
kc11 = 1;
KM12 = 72; %(-)-trans-Isopiperitenol dehydrogenase (IsoDH)
kc12 = 0.002;
KM13 = 1; %(-)-Isopiperitenone reductase (IsoR)
kc13 = 1.3;
KM14 = 270; %(+)-cis-Isopulegone isomerase (IsoI)
kc14 = 1;
KM15 = 10; %(+)-Menthofuran synthase (MFS)
kc15 = 0.9;
KM16a = 2.3; %(+)-Pulegone reductase (PR; product: (-)-menthone)
kc16a = 1.8;
KM16b = 2.3; %(+)-Pulegone reductase (PR; product: (+)-isomenthone)
kc16b = 1.8;
KM17a = 3; %(-)-Menthone:(-)-menthol reductase (MMR; substrate: (-)-menthone)
kc17a = 0.6;
KM17b = 41; %(-)-Menthone:(-)-menthol reductase (MMR; substrate: (+)-isomenthone)
kc17b = 0.6;
KM18af = 674; %(-)-Menthone:(+)-neomenthol reductase (MNR; substrate: (-)-menthone);
forward reaction)
kc18af = 0.06;
KM18ar = 1200; %(-)-Menthone:(+)-neomenthol reductase (MNR; substrate: (-)-menthone);
backward reaction)
kc18ar = 0.06;% estimated
KM18b = 1000; %(-)-Menthone:(+)-neomenthol reductase (MNR; substrate: (+)-isomenthone)
kc18b = 0.06;

Kic1=96; % Product inhibition constant (Geranyl diphosphate acting on IPPI)
Kic2=300; % Product inhibition constant ((+)-menthofuran acting on PR)
% Competitive inhibition mechanism

Kis=112; % Substrate Inhibition constant ((+)-pulegone acting on PR)
% Uncompetitive inhibition mechanism

z=100; % Factor to account for selective retention of (+)-menthofuran in
secretory cells

w=0.05; % Factor to account for the rapid excretion of (+)-pulegone from
secretory cells into oil storage cavity

% The model also takes into account that each enzyme shows a particular transient
% pattern of expression. This pattern is approximated by a Gauss function.

%First peak of activity:

%f(x) = Comp * a * exp(-(t-b).^2)/(2*(c)^2)

%where Comp = Factor to adjust for the volume density of the compartment in which a
particular enzyme is active [Dimensionless]
% a = Factor defining the height of the Gaussian peak for enzyme activity
[ Units of concentration ]

```

```

%      t      = Time [s]
%      b      = Factor defining the position of the center of the Gaussian peak for
%              enzyme activity [s]
%      c      = Factor defining the width of the Gaussian peak for enzyme activity at
%              half maximum [s]

b1=1296000; % Defines the position of the center of the Gaussian peak for enzyme
            % activity.
            % Relevant to the following enzyme activities: LS, L3H, IsoDH, IsoR, IsoI,
            % MFS, PR
c1=800000; % Defines the width of the Gaussian peak for enzyme activity at half
            % maximum.
            % Relevant to the following enzyme activities: LS, L3H, IsoDH, IsoR, IsoI,
            % MFS, PR

b5=1800000; % Defines the position of the center of the Gaussian peak for enzyme
            % activity.
            % Relevant to the following enzyme activities: MMR, MNR

c5=900000; % Defines the width of the Gaussian peak for enzyme activity at half
            % maximum.
            % Relevant to the following enzyme activities: MMR, MNR

E1=(0.139)*0.03*exp((-t-b1).^2)/(2*(c1)^2); % DXS
E2=(0.139)*0.0225*exp((-t-b1).^2)/(2*(c1)^2); % DXR
E3=(0.139)*0.5*exp((-t-b1).^2)/(2*(c1)^2); % MCT
E4=(0.139)*0.0225*exp((-t-b1).^2)/(2*(c1)^2); % CMK
E5=(0.139)*0.5*exp((-t-b1).^2)/(2*(c1)^2); % MECPS
E6=(0.139)*0.5*exp((-t-b1).^2)/(2*(c1)^2); % HDS
E7a=(0.139)*0.2*exp((-t-b1).^2)/(2*(c1)^2); % HDR (product: DMAPP)
E7b=(0.139)*0.04*exp((-t-b1).^2)/(2*(c1)^2); % HDR (product: IPP)
E8=(0.139)*0.3*exp((-t-b1).^2)/(2*(c1)^2); % IPPI
E9=(0.139)*0.1*exp((-t-b1).^2)/(2*(c1)^2); % GPPS
E10=(0.139)*0.017*exp((-t-b1).^2)/(2*(c1)^2); % LS
E11=(0.365)*0.003*exp((-t-b1).^2)/(2*(c1)^2); % L3H
E12=(0.044)*10*exp((-t-b1).^2)/(2*(c1)^2); % IsoDH
E13=(0.204)*0.34*exp((-t-b1).^2)/(2*(c1)^2); % IsoR
E14=(0.204)*0.34*exp((-t-b1).^2)/(2*(c1)^2); % IsoI
E15=(0.365)*0.00007*exp((-t-b1).^2)/(2*(c1)^2); % MFS
E16a=(0.204)*0.0015*exp((-t-b1).^2)/(2*(c1)^2); % PR (product: (-)-menthone)
E16b=(0.204)*0.00015*exp((-t-b1).^2)/(2*(c1)^2); % PR (product: (+)-isomenthone)
E17a=(0.204)*0.0011*exp((-t-b5).^2)/(2*(c5)^2); % MMR (product: (-)-menthol)
E17b=(0.204)*0.0011*exp((-t-b5).^2)/(2*(c5)^2); % MMR (product: (+)-neoisomenthol)
E18a=(0.204)*0.00001*exp((-t-b5).^2)/(2*(c5)^2); % MNR (product: (+)-neomenthol)
E18b=(0.204)*0.00001*exp((-t-b5).^2)/(2*(c5)^2); % MNR (product: (+)-isomenthol)

% The model also takes into account that the glandular trichome density (GN) changes over
% time. This behavior is approximated using a logistic function:

c=5*10^5; % parameter approximating slope of exponential phase of sigmoid curve
k=1/8*10^4; % parameter approximating shape of sigmoid curve

GN = 1+ 1/(1+c*exp(-k*t)); % at day 15, gland number is 86.7 % of total gland number at
                        % day 30

%Species Equations

if t < 1296000 % (patterns of enzymes from 0 to 15 days after leaf initiation)

xdot=[GN*(-(kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2)))); % Variation of
GAP
      GN*(-(kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2)))); % Variation of
Pyruvate (same expression as for GAP)
      GN*((kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2)))-((KM2r*kc2f*E2*x(3)-
KM2f*kc2r*E2*x(4))/(KM2f*KM2r+KM2r*x(3)+KM2f*x(4)))); % Variation of DOXP
      GN*((KM2r*kc2f*E2*x(3)-KM2f*kc2r*E2*x(4))/(KM2f*KM2r+KM2r*x(3)+KM2f*x(4)))-
(kc3*E3*x(4)/(x(4)+KM3)); % Variation of ME4P
      GN*((kc3*E3*x(4)/(x(4)+KM3))-(kc4*E4*x(5)/(x(5)+KM4))); % Variation of CDP-ME

```

```

GN*((kc4*E4*x(5)/(x(5)+KM4)-(kc5*E5*x(6)/(x(6)+KM5))); % Variation of CDP-
ME2P
GN*((kc5*E5*x(6)/(x(6)+KM5)-(kc6*E6*x(7)/(x(7)+KM6))); % Variation of MECPP
GN*((kc6*E6*x(7)/(x(7)+KM6)-(kc7*E7a*x(8)/(x(8)+KM7))-(kc7*E7b*x(8)/(x(8)+KM7)));
% Variation of HMB-PP
GN*((kc7*E7a*x(8)/(x(8)+KM7)+(kc8f*E8*x(10)/(x(10)+KM8f*(1+(x(11)/Kic1))))-
(kc8r*E8*x(9)/(x(9)+KM8r*(1+(x(11)/Kic1))))-
((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b)); %Variation of
DMAPP
GN*((kc7*E7b*x(8)/(x(8)+KM7)+(kc8r*E8*x(9)/(x(9)+KM8r*(1+(x(11)/Kic1))))-
(kc8f*E8*x(10)/(x(10)+KM8f*(1+(x(11)/Kic1))))-
((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b)); %Variation of
IPP
GN*((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b))-
(kc10*E10*x(11)/(x(11)+KM10)); %Variation of GPP
GN*((kc10*E10*x(11)/(x(11)+KM10)-(kc11*E11*x(12)/(x(12)+KM11))); % Variation of LIM
GN*((kc11*E11*x(12)/(x(12)+KM11)-(kc12*E12*x(13)/(x(13)+KM12))); % Variation of
IPPo1
GN*((kc12*E12*x(13)/(x(13)+KM12)-(kc13*E13*x(14)/(x(14)+KM13))); % Variation of
IPPone
GN*((kc13*E13*x(14)/(x(14)+KM13)-(kc14*E14*x(15)/(x(15)+KM14))); % Variation of
CIPUL
GN*((kc14*E14*x(15)/(x(15)+KM14))-
(kc16a*E16a*x(16)/(x(16)+KM16a*(1+z*(x(17))/Kic2)))-
(kc16b*E16b*x(16)/(x(16)+KM16b*(1+z*(x(17))/Kic2)))-
(w*kc16a*E16a*x(16)/(KM16a+x(16)*(1+x(16)/Kis)))-
(w*kc16b*E16b*x(16)/(KM16b+x(16)*(1+x(16)/Kis)))-(kc15*E15*x(16)/(x(16)+KM15)); %
Variation of PUL
GN*(kc15*E15*x(16)/(x(16)+KM15)); % Variation of MF

GN*((kc16b*E16b*x(16)/(x(16)+KM16b*(1+z*(x(17))/Kic2)))+(w*kc16b*E16b*x(16)/(KM16b+x(16)*
(1+x(16)/Kis)))-(kc17b*E17b*x(18)/(x(18)+KM17b))-(kc18b*E18b*x(18)/(x(18)+KM18b)); %
Variation of IMone

GN*((kc16a*E16a*x(16)/(x(16)+KM16a*(1+z*(x(17))/Kic2)))+(w*kc16a*E16a*x(16)/(KM16a+x(16)*
(1+x(16)/Kis)))-((KM18ar*kc18af*E18a*x(19)-
KM18af*kc18ar*E18a*x(20))/(KM18af*KM18ar+KM18ar*x(19)+KM18af*x(20)))-
(kc17a*E17a*x(19)/(x(19)+KM17a)); % Variation of Mone
GN*((KM18ar*kc18af*E18a*x(19)-
KM18af*kc18ar*E18a*x(20))/(KM18af*KM18ar+KM18ar*x(19)+KM18af*x(20)); % Variation of NMol
GN*(kc17a*E17a*x(19)/(x(19)+KM17a)); % Variation of Mol
GN*(kc18b*E18b*x(18)/(x(18)+KM18b)); % Variation of IMol
GN*(kc17b*E17b*x(18)/(x(18)+KM17b)); % Variation of
NIMol

else t>= 1296000 %(patterns of enzymes from 15 - 40 days after leaf initiation)

% Second peak of enzyme activity:

b2=1814400; % Defines the position of the center of the second Gaussian peak for
enzyme activity.
% Relevant to the following enzyme activities: PR
c2=1420000; % Defines the width of the second Gaussian peak for enzyme activity at
half maximum.
% Relevant to the following enzyme activities: PR

b4=2160000; % Defines the position of the center of the second Gaussian peak for
enzyme activity.
% Relevant to the following enzyme activities: IsoDH, IsoR, IsoI
c4=170000; % Defines the width of the second Gaussian peak for enzyme activity at
half maximum.
% Relevant to the following enzyme activities: IsoDH, IsoR, IsoI

E12=(0.044)*1*exp((-t-b4).^2)/(2*(c4)^2); % IsoDH
E13=(0.204)*0.0044*exp((-t-b2).^2)/(2*(c2)^2); % IsoR
E14=(0.204)*0.0044*exp((-t-b2).^2)/(2*(c2)^2); % IsoI
E16a=(0.204)*0.00014*exp((-t-b2).^2)/(2*(c2)^2); % PR (product: (-)-menthone)
E16b=(0.204)*0.000014*exp((-t-b2).^2)/(2*(c2)^2); % PR (product: (+)-isomenthone)

```

```

xdot=[GN*(-(kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2))); % Variation of
GAP
GN*(-(kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2))); % Variation of
Pyruvate (same expression as for GAP)
GN*((kclb*E1*x(2)*x(1)/(Kia*KM1b+KM1a*x(2)+KM1b*x(1)+x(1)*x(2)))-((KM2r*kc2f*E2*x(3)-
KM2f*kc2r*E2*x(4))/(KM2f*KM2r+KM2r*x(3)+KM2f*x(4))); % Variation of DOXP
GN*(((KM2r*kc2f*E2*x(3)-KM2f*kc2r*E2*x(4))/(KM2f*KM2r+KM2r*x(3)+KM2f*x(4)))-
(kc3*E3*x(4)/(x(4)+KM3))); % Variation of ME4P
GN*((kc3*E3*x(4)/(x(4)+KM3))-(kc4*E4*x(5)/(x(5)+KM4))); % Variation of CDP-ME
GN*((kc4*E4*x(5)/(x(5)+KM4))-(kc5*E5*x(6)/(x(6)+KM5))); % Variation of CDP-
ME2P
GN*((kc5*E5*x(6)/(x(6)+KM5))-(kc6*E6*x(7)/(x(7)+KM6))); % Variation of MEcPP
GN*((kc6*E6*x(7)/(x(7)+KM6))-(kc7*E7a*x(8)/(x(8)+KM7))-(kc7*E7b*x(8)/(x(8)+KM7)));
% Variation of HMB-PP
GN*((kc7*E7a*x(8)/(x(8)+KM7))+(kc8f*E8*x(10)/(x(10)+KM8f*(1+(x(11)/Kic1))))-
(kc8r*E8*x(9)/(x(9)+KM8r*(1+(x(11)/Kic1))))-
((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b)); %Variation of
DMAPP
GN*((kc7*E7b*x(8)/(x(8)+KM7))+(kc8r*E8*x(9)/(x(9)+KM8r*(1+(x(11)/Kic1))))-
(kc8f*E8*x(10)/(x(10)+KM8f*(1+(x(11)/Kic1))))-
((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b)); %Variation of
IPP
GN*(((kc9a*E9*KM9b*x(9)+kc9b*E9*KM9a*x(10))/(KM9b*x(9)+KM9a*x(10)+KM9a*KM9b))-
(kc10*E10*x(11)/(x(11)+KM10))); %Variation of GPP
GN*((kc10*E10*x(11)/(x(11)+KM10))-(kc11*E11*x(12)/(x(12)+KM11))); % Variation of LIM
GN*((kc11*E11*x(12)/(x(12)+KM11))-(kc12*E12*x(13)/(x(13)+KM12))); % Variation of
IPPol
GN*((kc12*E12*x(13)/(x(13)+KM12))-(kc13*E13*x(14)/(x(14)+KM13))); % Variation of
IPPone
GN*((kc13*E13*x(14)/(x(14)+KM13))-(kc14*E14*x(15)/(x(15)+KM14))); % Variation of
CIPUL
GN*((kc14*E14*x(15)/(x(15)+KM14))-
(kc16a*E16a*x(16)/(x(16)+KM16a*(1+z*(x(17)/Kic2)))-
(kc16b*E16b*x(16)/(x(16)+KM16b*(1+z*(x(17)/Kic2)))-
(w*kc16a*E16a*x(16)/(KM16a+x(16)*(1+x(16)/Kis)))-
(w*kc16b*E16b*x(16)/(KM16b+x(16)*(1+x(16)/Kis)))-
(kc15*E15*x(16)/(x(16)+KM15))); %
Variation of PUL
GN*(kc15*E15*x(16)/(x(16)+KM15)); % Variation of MF
GN*((kc16b*E16b*x(16)/(x(16)+KM16b*(1+z*(x(17)/Kic2)))+(w*kc16b*E16b*x(16)/(KM16b+x(16)*
(1+x(16)/Kis)))-
(kc17b*E17b*x(18)/(x(18)+KM17b))-(kc18b*E18b*x(18)/(x(18)+KM18b))); %
Variation of IMone
GN*((kc16a*E16a*x(16)/(x(16)+KM16a*(1+z*(x(17)/Kic2)))+(w*kc16a*E16a*x(16)/(KM16a+x(16)*
(1+x(16)/Kis)))-
((KM18ar*kc18af*E18a*x(19)-
KM18af*kc18ar*E18a*x(20))/(KM18af*KM18ar+KM18ar*x(19)+KM18af*x(20))-
(kc17a*E17a*x(19)/(x(19)+KM17a))); % Variation of Mone
GN*((KM18ar*kc18af*E18a*x(19)-
KM18af*kc18ar*E18a*x(20))/(KM18af*KM18ar+KM18ar*x(19)+KM18af*x(20))); % Variation of NMol
GN*(kc17a*E17a*x(19)/(x(19)+KM17a)); % Variation of Mol
GN*(kc18b*E18b*x(18)/(x(18)+KM18b)); % Variation of IMol
GN*(kc17b*E17b*x(18)/(x(18)+KM17b)); % Variation of NIMol

```

end

## CHAPTER 5

### Conclusions

The work presented in this dissertation reports on novel approaches to unravel the complex regulation of monoterpene biosynthesis in peppermint.

Based upon a vast amount of physiological and biochemical data (literature and our own experiments), we developed a kinetic mathematical model that accurately describes the behavior of the peppermint monoterpene biosynthetic pathway under various experimental conditions. Model simulations were used to generate non-trivial, testable hypotheses regarding poorly understood regulatory mechanisms, and modeling-guided follow-up experiments were used to demonstrate an as yet unidentified role for (+)-menthofuran as a competitive inhibitor of PR. A similar integrative approach of iterative modeling and experimental testing identified glandular trichome distribution as the primary factor determining oil yield.

After model improvements to include glandular trichome density and biosynthetic gene expression patterns, a second generation model accurately simulated the dynamics of oil accumulation in transgenic and wild-type plants grown under normal and adverse environmental conditions. These simulations indicated that we can account for all major factors affecting essential oil yield and composition in wild-type plants and selected transgenic lines. The experimental results outlined in this dissertation illustrate the utility of integrative approaches of mathematical modeling and experimental testing as a powerful tool that can be employed to guide metabolic engineering efforts aimed at modulating monoterpenoid essential oil profiles in peppermint leaves.

## Future Directions

The work represented in this dissertation has demonstrated the utility of combining mathematical modeling and experimental testing in an iterative cycle to unravel the regulation of monoterpene biosynthesis in peppermint. An expansion of this approach and further model refinements will be briefly discussed here.

The established kinetic mathematical model should be expanded to include all enzymatic steps from sugar (most likely raffinose) import, via glycolysis and the pentose phosphate pathway, to monoterpene production.

Additional modeling efforts could include thermodynamically constrained models. Monoterpene biosynthesis requires a series of redox transformations. Therefore, knowing the redox status of the specialized secretory cells would facilitate the extension of model to include thermodynamic constraints, such as the thermodynamically feasible ratio ATP/ADP and/or NADP/NADPH. In spite of the limited availability of thermodynamic data, group contribution methodologies have increasingly been used to estimate thermodynamic properties in living systems (Mavrovouniotis, 1990; Mavrovouniotis 1991). In fact, several rigorous models of the metabolic chemistry of a variety of microorganisms have been already developed and the good of those models can eventually be extended to monoterpene metabolism modeling improvements (Reed et al., 2003; Price et al., 2004).

Finally, knowing that monoterpene biosynthesis takes place in different cellular compartments, modeling improvements should include additional constraints to account

for diffusive or facilitated transport of intermediate metabolites between compartments. (Sanders, 1990; Kehres, 1992; Borstlap and Schuurmans, 2004) This would greatly improve the accuracy of the model, account for additional potential variables, and would be another step toward a fully predictive mathematical model.

## References

Mavrovouniotis, M. L. 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.* 36:1070–1082.

Mavrovouniotis, M. L. 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266:14440–14445.

Price, N. D., J. L. Reed, and B. O. Palsson. 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2:886–897.

Reed, J. L., T. D. Vo, C. H. Schilling, and B. O. Palsson. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:54.1–54.12.

Sanders D. 1990 Kinetic modeling of plant and fungal membrane transport. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 1990. 41:77-107.

Kehres D.G. 1992. A kinetic model for binding protein-mediated arabinose transport *Protein Science, I*, 1661-1665.

Borstlap, J.A and Schuurmans M.J. 2004. Sucrose Transport into Plasma Membrane Vesicles from Tobacco Leaves by H<sup>+</sup> Symport or Counter Exchange Does not Display a Linear Component. *Journal of Membrane Biology* 198(1):31-42.

## **APPENDIX**

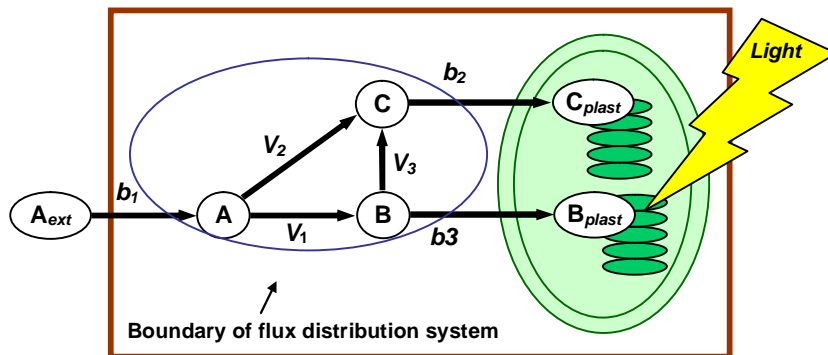


**A. SUPPLEMENTARY MATERIAL FOR THE REVIEW PAPER  
ENTITLED: EXPERIMENTAL AND MATHEMATICAL  
APPROACHES TO MODELING PLANT METABOLIC NETWORKS**

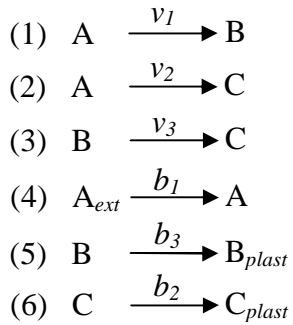
**A. SOLVING A LINEAR OPTIMIZATION PROBLEM IN FLUX BALANCE  
ANALYSIS**

Let us consider the metabolic network defined in Scheme 1. A plant cell imports the metabolite  $A_{ext}$  from the apoplastic space (the internal metabolite is referred to as  $A$ ). The metabolic network consists of a series of reactions that result in the formation of metabolites  $B$  and  $C$ . These cytosolically synthesized metabolites are then transported to plastids (they are now referred to as  $B_{plast}$  and  $C_{plast}$ ) and associate with thylakoids. Our purpose is to determine the internal flux distribution, using FBA, in the subnetwork circumscribed by the blue oval of Scheme 1.

**Scheme 1.** Example network to illustrate flux balance analysis.



The metabolic reactions and transport processes can be summarized as follows ( $v_i$  are reaction rates,  $b_i$  are transport rates):



The mass balance of a metabolite (variation over time  $dM/dt$ ) is defined as the difference between the rate(s) of production and the rate(s) of consumption. Applied to all metabolites in our network, the following set of coupled ordinary differential equations can be formulated:

**Scheme 2.** Dynamic mass balance equations.

$$\frac{dA}{dt} = b_1 - v_1 - v_2$$

$$\frac{dB}{dt} = v_1 - b_3 - v_3$$

$$\frac{dC}{dt} = v_2 + v_3 - b_2$$

We have now defined the metabolic network and have derived the corresponding dynamic mass balance equations, which can also be rewritten in matrix notation (Scheme 3). To calculate the flux distribution in the network under consideration, we need to identify imposed restrictions (constraints) to the system. If we consider that the macroscopic variables (metabolite concentrations and fluxes) do not change considerably

over a certain time span (steady-state assumption), the dynamic mass balance equations can be represented as  $S \cdot \mathbf{v} = \mathbf{0}$ , where  $S$  is the stoichiometric matrix and  $\mathbf{v}$  the flux vector:

**Scheme 3.** Mass balance equations in matrix form.

$$\frac{d}{dt} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The arrows in Scheme 3 indicate which flux in the flux vector corresponds to which stoichiometric coefficient in the first line of the stoichiometric matrix based on the first equation (specifying  $dA/dt$ ) in Scheme 2. Lines 2 and 3 of the stoichiometric matrix are obtained in the same way for  $dB/dt$  and  $dC/dt$ , respectively. For our example the number of unknown fluxes ( $v_1$ ,  $v_2$ ,  $v_3$ ,  $b_1$ ,  $b_2$ , and  $b_3$ ) is greater than the number of metabolites ( $A$ ,  $B$ , and  $C$ ), which means that the steady-state solution of fluxes is under-determined. Thus, additional constraints such as measurable metabolite concentrations and measurable fluxes can be used to uniquely determine the flux distribution. Let us assume that we have measured the concentrations of metabolites  $A_{ext}$  and  $A$  over a certain period of time; we then calculate the uptake rate  $b_1$  (for this example we assume this rate is in the range of up to  $10 \text{ pmol (g fresh weight} \cdot \text{s)}^{-1}$ ). Let us also assume that we have measured the concentrations of metabolites  $B_{plast}$  and  $C_{plast}$  and we observed that their ratio under various conditions is always greater than or equal to  $3 : 1$ . Further assumptions are that the cell under consideration is experiencing high light conditions

and that compound  $B_{plast}$  is an essential component of the high light response of plants, which should be synthesized at high levels under the experimental conditions. In order to determine the internal flux distribution, we can thus formulate an objective function  $Z$  (the cell maximizes the production of  $B_{plast}$  and, because  $B_{plast}$  and  $C_{plast}$  are linked as defined above, also  $C_{plast}$ ), which can be treated as a linear optimization problem:

$$\text{Maximize } Z = B_{plast} + C_{plast} = b_2 + b_3 \quad \text{Objective function}$$

Summary of constraints:

- (1)  $S \cdot v = 0$  *Mass balance constraint assuming steady-state*
- (2)  $b_1 \leq 10$  *Inferred by measurement of concentrations of  $A_{ext}$  and  $A$*
- (3)  $b_3 \geq 3 \cdot b_2$  *Inferred by measurement of concentrations of  $B_{plast}$  and  $C_{plast}$*
- (4)  $v_1, v_2, v_3, b_1, b_2, b_3 \geq 0$  *Assumption that all reactions and the direction of transport are not reversible*

Optimization problems with an objective function are oftentimes solved using an approach called linear programming (LP). This method identifies one solution (in this case a maximum) that satisfies all constraints. One popular technique for the numerical solution of an LP problem is the simplex algorithm. The simplex method uses an iterative process in which the objective function continuously increases until optimality is reached. Linear Programming methods (LP) identify an extreme in the feasible region (in this case a maximum) which satisfies the set of constraints. For the analytical solution of the LP problem, the simplex method will be used. (For a detailed explanation about the simplex method algorithm see, Wayne L. Winston et. al., (2003) "Introduction to Mathematical Programming", 4th edition, ed. Duxbury).

**Step1: Defining a system of linear equations.**

The simplex method proceeds from one extreme point in the feasible region to another, in such a way that the objective function continuously increases until optimality is reached. LP problems must be converted into augmented form (slack form) before being solved by the simplex algorithm. This means that all  $\leq$  and  $\geq$  constraints (non-equalities) have to be converted to equalities, which is enabled by the introduction of what is called a slack variable  $s_i$  (note that we currently do not know the algebraic sign of  $s_i$ , so we have to consider both  $+s_i$  and  $-s_i$  for converting the mass balance equations; the exception is  $s_5$ , the algebraic sign of which is positive based on *constraint (2)*). The objective function  $Z$  is first set to zero and subsequent iterations are used to maximize its value. Following is the resulting system of linear equations:

**Scheme 4**

$Z -b_2 -b_3$	$= 0$	<i>Derived from objective function</i>
$\frac{dA}{dt} = -v_1 -v_2 +b_1 +s_1$	$= 0$	} <i>Mass balance equations with positive slack variable</i>
$\frac{dB}{dt} = +v_1 -v_3 -b_3 +s_2$	$= 0$	
$\frac{dC}{dt} = +v_2 +v_3 -b_2 +s_3$	$= 0$	
$b_3 = 3 b_2 +s_4 \quad \text{or} \quad b_3 -3 b_2 -s_4$	$= 0$	<i>Derived from constraint (3)</i>
$b_1 + s_5$	$= 10$	<i>Derived from constraint (2)</i>
$\frac{dA}{dt} = -v_1 -v_2 +b_1 -s_6$	$= 0$	} <i>Mass balance equations with negative slack variable</i>
$\frac{dB}{dt} = +v_1 -v_3 -b_3 -s_7$	$= 0$	
$\frac{dC}{dt} = +v_2 +v_3 -b_2 -s_8$	$= 0$	

**Step 2: Setting up the initial tableau.**

The simplex method is a tabular solution algorithm. Each tableau represents a movement from one extreme to another, making sure that the objective function increases iteration after iteration until the optimal solution is obtained. The tableau contains the set of algebraic equations, including the optimization function ( $Z$ ), the system variables ( $v_1, v_2, v_3$ ) and slack variables ( $s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8$ ) and the so called, right hand side vector (RHS) (right side of system equation). The difference between the number of variables and the number of equations specifies the degrees of freedom associated with the problem. Any solution will thus include a number of variables of arbitrary value and the simplex algorithm uses zero as its arbitrary value (which means that we have to set 5 variables to zero, one at a time). The variables set to zero are called non-basic variables, whereas the variables not set to zero are called basic variables. For the initial tableau, the basic variables are always the slack variables, which results in the following tableau (constructed based on the order of equations shown in Scheme 4):

**Scheme 5**

Eq./row No.	Basic variable	Coefficients															RHS	
		Z	$v_1$	$v_2$	$v_3$	$b_1$	$b_2$	$b_3$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$		
1	Z	1	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0
2	$s_1$	0	-1	-1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
3	$s_2$	0	1	0	-1	0	0	-1	0	1	0	0	0	0	0	0	0	0
4	$s_3$	0	0	1	1	0	-1	0	0	0	1	0	0	0	0	0	0	0
5	$s_4$	0	0	0	0	0	-3	1	0	0	0	-1	0	0	0	0	0	0
6	$s_5$	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	10	0
7	$s_6$	0	-1	-1	0	1	0	0	0	0	0	0	0	-1	0	0	0	0
8	$s_7$	0	1	0	-1	0	0	-1	0	0	0	0	0	0	-1	0	0	0
9	$s_8$	0	0	1	1	0	-1	0	0	0	0	0	0	0	0	-1	0	0

The ‘values’ correspond to the numbers on the right side of the equal sign in the equations of Scheme 4). The tableau represents the solution  $v_1, v_2, v_3, b_1, b_2, b_3 = 0$  (we

know this because these are non-basic variables),  $s_1, s_2, s_3, s_4, s_6, s_7, s_8 = 0$ , and  $s_5 = 10$  (the values of the last column correspond to the basic variables in the first column; the numbers in all other columns are stoichiometric coefficients, not values). A tableau represents an optimal solution if (1) all ‘values’ in the last column are non-negative (feasibility condition), (2) the objective row contains no negative entries in the columns of non-basic variables (optimality condition), (3) the objective row contains zero entries in the columns of basic variables, and (4) all imposed constraints are satisfied ( $b_3 = 3b_2$ ;  $b_1 \leq 10$ ;  $v_1, v_2, v_3, b_1, b_2, b_3 \geq 0$ ; mass balance equations for A, B, C in Scheme 2). This is obviously not the case in our example and an optimization is thus needed. This optimization process uses the simplex algorithm for solving LP problems, which is explained using a simple case study at [http://en.wikipedia.org/wiki/Simplex\\_algorithm\\_method](http://en.wikipedia.org/wiki/Simplex_algorithm_method). First, we multiply both sides of equations (4), (5), and (9) by (-1):

Eq./row No.	Basic variable	Coefficients															RHS		
		Z	$v_1$	$v_2$	$v_3$	$b_1$	$b_2$	$b_3$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$			
1	Z	1	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	
2	$s_1$	0	-1	-1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
3	$s_2$	0	1	0	-1	0	0	-1	0	1	0	0	0	0	0	0	0	0	0
4	$s_3$	0	0	-1	-1	0	1	0	0	0	-1	0	0	0	0	0	0	0	0
5	$s_4$	0	0	0	0	0	3	-1	0	0	0	1	0	0	0	0	0	0	0
6	$s_5$	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	10	0
7	$s_6$	0	-1	-1	0	1	0	0	0	0	0	0	0	-1	0	0	0	0	0
8	$s_7$	0	1	0	-1	0	0	-1	0	0	0	0	0	0	-1	0	0	0	0
9	$s_8$	0	0	-1	-1	0	1	0	0	0	0	0	0	0	0	1	0	0	0

**Step 3: Select pivot column and pivot.**

To identify the entering variable, select the negative number with the largest magnitude in row 1 (Z equation), excluding the last column (RHS). The column under this number is called the ‘pivot column’ (in our example there are two columns with the

value ‘-1’ (those under b2 and b3)). The pivot row is generally determined as the smallest non-negative ratio between a value in the candidate pivot columns and the corresponding value in the RHS column. In our example, most RHS values are ‘0’, so we just select a row with a positive value as pivot row (equation 9) and obtain the pivot as the intersection of pivot column and pivot row (red circle).

**Tableau 1**

Eq./row No.	Basic variable	Coefficients															RHS	
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>		
1	Z	1	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0
2	s <sub>1</sub>	0	-1	-1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
3	s <sub>2</sub>	0	1	0	-1	0	0	-1	0	1	0	0	0	0	0	0	0	0
4	s <sub>3</sub>	0	0	-1	-1	0	1	0	0	0	-1	0	0	0	0	0	0	0
5	s <sub>4</sub>	0	0	0	0	0	3	-1	0	0	0	1	0	0	0	0	0	0
6	s <sub>5</sub>	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	10
7	s <sub>6</sub>	0	-1	-1	0	1	0	0	0	0	0	0	0	-1	0	0	0	0
8	s <sub>7</sub>	0	1	0	-1	0	0	-1	0	0	0	0	0	0	0	-1	0	0
9	s <sub>8</sub>	0	0	-1	-1	0	1	0	0	0	0	0	0	0	0	0	1	0

**Step 3: Obtain next tableau.**

In the next tableau, we must obtain zero in the pivot column for the first row (Z equation). The row operation is as follows:

$$\text{New row} = \text{old row} - \left\{ \text{number in the old row that belong to the pivot column} \right. \\ \left. \text{multiplied by the pivot row divided by the pivot} \right\}$$

The row calculations must be performed for all the remaining numbers of the pivot column, so its elements are now zeros, except for the pivot, which is one. The resulting tableau must be evaluated to determine if it contains the optimal solution. If all RHS values are non-negative (known as the feasibility condition) and if all elements of the first



row (Z row) are non-positive (optimality condition) the tableau contains the optimal solution. If not, a new iteration is required and the steps 3 and 4 must be repeated until the optimal solution is reached. To reach the optimal solution, seven iterations were required. Following are the complete tableaus:

**Tableau 2**

Eq./row No.	Basic variable	Coefficients															
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	RHS
1	Z	1	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	1	0
2	s <sub>1</sub>	0	-1	-1	0	1	0	0	1	0	0	0	0	0	0	0	0
3	s <sub>2</sub>	0	1	0	-1	0	0	-1	0	1	0	0	0	0	0	0	0
4	s <sub>3</sub>	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0
5	s <sub>4</sub>	0	0	3	3	0	0	-1	0	0	0	1	0	0	0	-3	0
6	s <sub>5</sub>	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	10
7	s <sub>6</sub>	0	1	1	0	-1	0	0	0	0	0	0	0	1	0	0	0
8	s <sub>7</sub>	0	-1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
9	b <sub>2</sub>	0	0	-1	-1	0	1	0	0	0	0	0	0	0	0	1	0

In our next iteration  $v_2$  was selected as the entering variable. Both sides of equations (7) and (8) were multiplied by (-1) and a new pivot column, row and pivot were obtained:

**Tableau 3**

Eq./row No.	Basic variable	Coefficients															
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	RHS
1	Z	1	1	0	-1	-1	0	-1	0	0	0	0	0	1	0	1	0
2	s <sub>1</sub>	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
3	s <sub>2</sub>	0	1	0	-1	0	0	-1	0	1	0	0	0	0	0	0	0
4	s <sub>3</sub>	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0
5	s <sub>4</sub>	0	-3	0	3	3	0	-1	0	0	0	1	0	-3	0	-3	0
6	s <sub>5</sub>	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	10
7	v <sub>2</sub>	0	1	1	0	-1	0	0	0	0	0	0	0	1	0	0	0
8	s <sub>7</sub>	0	-1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
9	b <sub>2</sub>	0	1	0	-1	-1	1	0	0	0	0	0	0	1	0	1	0

In the next iteration  $v_3$  was chosen as the entering variable:

**Tableau 4**

Eq./row No.	Basic variable	Coefficients															RHS
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	
1	Z	1	0	0	0	-1	0	0	0	0	0	0	0	1	1	1	0
2	S1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
3	S2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
4	S3	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0
5	S4	0	0	0	0	3	0	-4	0	0	0	1	0	-3	-3	-3	0
6	S5	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	10
7	V2	0	1	1	0	-1	0	0	0	0	0	0	1	0	0	0	0
8	V3	0	-1	0	1	0	0	1	0	0	0	0	0	1	0	0	0
9	B2	0	0	0	0	-1	1	1	0	0	0	0	1	1	1	0	0

Next, **b<sub>1</sub>** was selected as entering variable. Equation (5) was divided by 3 to obtain the next tableau:

**Tableau 5**

Eq./row No.	Basic variable	Coefficients															RHS
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	
1	Z	1	0	0	0	0	0	-1	0	0	0	0.3	0	0	0	0	0
2	s1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
3	s2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
4	s3	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0
5	b1	0	0	0	0	1	0	-1	0	0	0	0.3	0	-1	-1	-1	0
6	s5	0	0	0	0	0	0	1.3	0	0	0	-0	1	1	1	1	10
7	v2	0	1	1	0	0	0	-1	0	0	0	0.3	0	0	-1	-1	0
8	v3	0	-1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
9	b2	0	0	0	0	0	1	-0	0	0	0	0.3	0	0	0	0	0

Next, **b<sub>3</sub>** was selected as entering variable:

**Tableau 6**

Eq./row No.	Basic variable	Coefficients															RHS
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	
1	Z	1	-1.3	0	1.3	0	0	0	0	0	0	0.3	0	0	1	0	0
2	s1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
3	s2	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
4	s3	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	-1	0
5	b1	0	-1.3	0	1.3	1	0	0	0	0	0	0.3	0	-1	0	-1	0
6	s5	0	1.3	0	-1.3	0	0	0	0	0	0	-0	1	1	-0	1	10
7	v2	0	-0.3	1	1.3	0	0	0	0	0	0	0.3	0	0	0	-1	0
8	b3	0	-1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
9	b2	0	-0.3	0	0.3	0	1	0	0	0	0	0.3	0	0	0	0	0

Finally, dividing equation (6) by 1.3, we identified the last pivot number and obtained the optimal solution:

**Tableau 7**

Eq./row No.	Basic variable	Coefficients															
		Z	v <sub>1</sub>	v <sub>2</sub>	v <sub>3</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	RHS
1	Z	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	10
2	s <sub>1</sub>	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
3	s <sub>2</sub>	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
4	s <sub>3</sub>	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0
5	b <sub>1</sub>	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	10
6	v <sub>1</sub>	0	1	0	-1	0	0	0	0	0	0	-0	1	1	-0	1	8
7	v <sub>2</sub>	0	0	1	1	0	0	0	0	0	0	0.3	0	0	0	-1	2
8	b <sub>3</sub>	0	0	0	0	0	0	1	0	0	0	-0	1	1	1	1	8
9	b <sub>2</sub>	0	0	0	0	0	1	0	0	0	0	0.3	0	0	0	0	2

**Optimal solution**

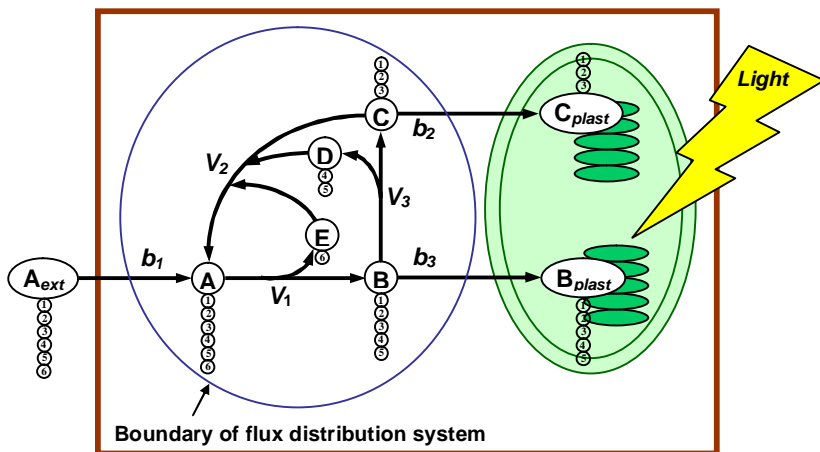
The following optimal solution for the problem under consideration (fluxes are expressed in pmol (g fresh weight • s)<sup>-1</sup>) was obtained:

$$b_1 = 10; b_2 = 2.5; b_3 = 7.5; v_1 = 7.5; v_2 = 2.5; v_3 = 0$$

**B. THE ISOTOPOMER BALANCING APPROACH**

The following metabolic network in considered:

**Scheme 6.** (Note: numbering of schemes is different from main text)



Isotopomer mapping matrices (IMMs) allow for expressing all isotopomer mass balances of a metabolite pool in a single matrix equation (Jeffrey et al., 1991; Künneke et al., 1993; Schmidt et al., 1997). In analogy to label distribution vectors of the AMM approach, IMMs use isotopomer distribution vectors (IDVs). Its notation is binary (hence the subscript ‘bin’) using a “0” for absence and a “1” for presence of isotope label. For metabolite *A* in the metabolic network depicted in Scheme 6, the IDV  $I_A$  will contain  $2^6 = 64$  elements ( $2^n$  isotopomers, where *n* is the number of carbons in the metabolite). The individual matrices shown in Scheme 9 are combined into one large matrix:

**Scheme 7.** (Note: numbering of schemes is different from main text)

$$I_A = \begin{bmatrix} I_A(1) \\ I_A(2) \\ I_A(3) \\ I_A(4) \\ \dots \\ I_A(64) \end{bmatrix} = \begin{bmatrix} I_A(000000_{bin}) \\ I_A(000001_{bin}) \\ I_A(000010_{bin}) \\ I_A(000011_{bin}) \\ \dots \\ I_A(111111_{bin}) \end{bmatrix}$$

The IDVs of product molecules are obtained by matrix multiplication of IMMs (which inform about the reaction mechanism) and IDVs of the reactants (Schmidt et. al., 1997). IMMs contain all pairs of reactant isotopomers responsible for the synthesis of specific product isotopomers in all positions of the product IDV. Thus, IMMs allow us to identify the substrate isotopomer from which a specific product isotopomer was synthesized. There will be as much IMMs as the number of pairs of reactants and product molecules in a biochemical reaction (Schmidt et. al., 1997). IMMs can be generated from AMMs in an iterative process, which we will illustrate for reaction (6) of Scheme 6.

**Scheme 8.** Construction of atom mapping matrices.

(Note: numbering of schemes is different from main text)

<b>Reaction</b>	<b>Function / Notation</b>
(1) $\longrightarrow A$	<i>Transport</i>
(2) $\xrightarrow{Enz1} B + E$	$[A > B]_{Enz1}; [A > E]_{Enz1}$
(3) $\longrightarrow B_{plast}$	<i>Transport</i>
(4) $B \xrightarrow{Enz3} C + D$	$[B > C]_{Enz3}; [B > D]_{Enz3}$
(5) $C \longrightarrow C_{plast}$	<i>Transport</i>
(6) $C + D + E \xrightarrow{Enz2} A$	$[C > A]_{Enz2}; [D > A]_{Enz2}; [E > A]_{Enz2}$

In order to obtain all possible labeling patterns of *A*, we have to consider all labeling possibilities in each one of the 2<sup>n</sup> isotopomers for the substrates involved in the reaction generating *A*.

First Iteration: We will define the label distribution according to the labeling pattern (000<sub>bin</sub>), (00<sub>bin</sub>) and (0<sub>bin</sub>) for the substrates *C*, *D*, and *E*, respectively. As before, *C*, *D* and *E* are the label distribution vectors containing a determined fractional enrichment;  $AMM_{C>A}$ ,  $AMM_{D>A}$  and  $AMM_{E>A}$  are the atom mapping matrices describing the transfer of carbon atoms from *C* to *A*, *D* to *A* and *E* to *A*, respectively; and *A* is the product isotopomer:

$$\begin{aligned}
 & \text{AMM}_{C>A} \bullet C + \text{AMM}_{D>A} \bullet D + \text{AMM}_{E>A} \bullet E = A \\
 & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \cdot [0] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
 \end{aligned}$$

The labeling pattern for  $A$  will be  $(000000_{\text{bin}})$  which corresponds to the index “1” in the IDV. Thus, we can assign a “1” to the first row of the first column in the  $\text{IMM}_{C>A}$ .

Second iteration: The labeling pattern of substrates  $C$ ,  $D$ , and  $E$  is assumed to be  $(000_{\text{bin}})$ ,  $(00_{\text{bin}})$  and  $(1_{\text{bin}})$ , respectively:

$$\begin{aligned}
 & \text{AMM}_{C>A} \bullet C + \text{AMM}_{D>A} \bullet D + \text{AMM}_{E>A} \bullet E = A \\
 & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \cdot [1] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}
 \end{aligned}$$

The isotopomer  $(000001_{\text{bin}})$  corresponds to the index “2” of the IDV of  $A$ ; thus, a “1” is assigned to the second row of the first column in the  $\text{IMM}_{C>A}$ .

Third Iteration: The labeling pattern of substrates  $C$ ,  $D$ , and  $E$  is assumed to be  $(000_{\text{bin}})$ ,  $(01_{\text{bin}})$  and  $(0_{\text{bin}})$ , respectively:

$$\begin{aligned}
 & \text{AMM}_{C>A} \bullet C + \text{AMM}_{D>A} \bullet D + \text{AMM}_{E>A} \bullet E = A \\
 & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \cdot [0] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}
 \end{aligned}$$

The isotopomer (000010<sub>bin</sub>) corresponds to the index “3” of the IDV of A; thus, a “1” is placed in the third row of the second column of the IMM<sub>C>A</sub>.

Fourth Iteration: The labeling pattern of *C*, *D*, and *E* is assumed to be (000<sub>bin</sub>), (01<sub>bin</sub>) and (1<sub>bin</sub>), respectively. Analogous to the process described above, the isotopomer (000011<sub>bin</sub>), which corresponds to the index 4 of the IDV, is obtained and a “1” can be assigned to the fourth row of the second column. This iterative process is repeated until all possible isotopomer distributions have been calculated.

When the IMMs for each pair of substrate and product molecules have been obtained, the product of these matrices and the corresponding IDVs of substrate molecules allows for the IDVs for product molecules to be calculated. For the reaction under consideration, the product IDV ( $I_A$ ) is calculated from the substrate IDVs ( $I_C$ ,  $I_D$ ,  $I_E$ ) and the corresponding IMMs (IMM<sub>C>A</sub>, IMM<sub>D>A</sub>, IMM<sub>E>A</sub>) (“ $\otimes$ ” specifies the element-wise multiplication of two column vectors with the same number of elements):

$$I_A = [\text{IMM}_{C>A} \bullet I_C] \otimes [\text{IMM}_{D>A} \bullet I_D] \otimes [\text{IMM}_{E>A} \bullet I_E]$$

In analogy to the AMM approach, the isotopomer distribution for the entire metabolic network of Scheme 6 is obtained by including the rate of production and the rate of

consumption of isotopomers for each metabolite, which is dependent of both metabolic ( $v_1, v_2, v_3$ ) and transport ( $b_1, b_2, b_3$ ) fluxes:

$$\frac{dI_A}{dt} = b_1 \cdot I_A + v_2 [IMM_{C>A} \cdot I_C] \otimes [IMM_{D>A} \cdot I_D] \otimes [IMM_{E>A} \cdot I_E] - v_1 \cdot I_A$$

$$\frac{dI_B}{dt} = v_1 [IMM_{A>B} \cdot I_A] - b_3 \cdot I_B - v_3 \cdot I_B$$

$$\frac{dI_C}{dt} = v_3 [IMM_{B>C} \cdot I_B] - b_2 \cdot I_C - v_2 \cdot I_C$$

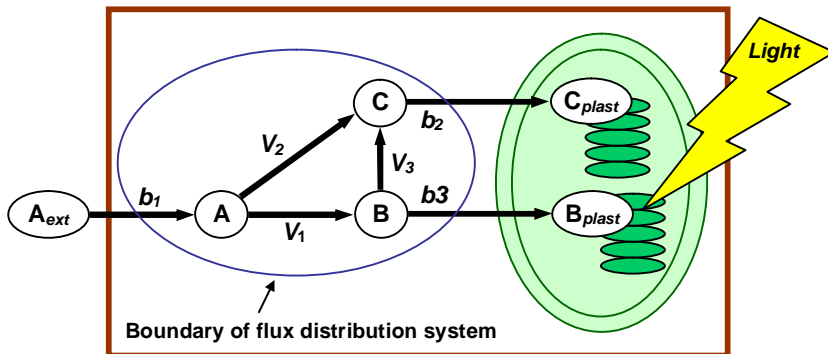
For a detailed explanation about the analytical solution and the interpretation of the results, see Schmidt et al. (1999).

### C. MATLAB® CODE FOR A SIMPLE KINETIC MODEL

We will consider the metabolic network described in Scheme 1 under the assumption that the transport fluxes  $b_1, b_2$  and  $b_3$  exert negligible control over flux; this means that only the irreversible reaction fluxes  $v_1, v_2$  and  $v_3$  need to be considered.

**Scheme 9.** Example network to illustrate kinetic modeling.

(Note: numbering of schemes is different from main text)





To obtain the mass balances for metabolites  $A$ ,  $B$  and  $C$ , the reaction network stoichiometry (I) is combined with Michaelis-Menten type enzyme kinetics (II) for each enzymatic step:

(I) Stoichiometric balances

$$\frac{dA}{dt} = -v_1 - v_2$$

$$\frac{dB}{dt} = v_1 - v_3$$

$$\frac{dC}{dt} = v_3 + v_2$$

(II) Michaelis-Menten rate constants

$$v_1 = v_{max(Enz1)} \frac{A}{A + K_m(Enz1)}$$

$$v_3 = v_{max(Enz3)} \frac{A}{A + K_m(Enz3)}$$

$$v_2 = v_{max(Enz2)} \frac{A}{A + K_m(Enz2)}$$

The variation of metabolite concentration over time ( $dM/dt$ ) is equal to the difference of its rate of formation and its rate of consumption. Let us assume that we have measured an initial substrate concentration ( $[A] = 1 \mu\text{M}$ ) and that we have obtained the following values for  $K_m$  and  $V_{max}$ :  $K_{m(Enz1)} = 0.8 \mu\text{M}$ ;  $K_{m(Enz2)} = 0.1 \mu\text{M}$ ;  $K_{m(Enz3)} = 0.004 \mu\text{M}$ ;  $V_{max(Enz1)} = 0.0018 \mu\text{M} \cdot \text{sec}^{-1}$ ;  $V_{max(Enz2)} = 0.0018 \mu\text{M} \cdot \text{sec}^{-1}$ ;  $V_{max(Enz3)} = 0.000012 \mu\text{M} \cdot \text{sec}^{-1}$ . For the simultaneous solution of the system of ODEs we used the 'ode45' function in the MATLAB® software:

```
%clear
xdot = zeros(3,1);
tspan = [0:0.1:2];
%vector for initial conditions
xdot0 = [1; 0; 0];
%Solving the ODE
[t,x] = ode45('kin', tspan, xdot0, []);
A=x(:,1);
B=x(:,2);
C=x(:,3);
plot(t,A,'b',t,B,'g',t,C,'r')
Legend('A','B','C')
```

```

%The kin function :
function xdot = kin(t,x,flag,E)
%[A]=x(1), [B]=x(2), [C]=x(3)
%Michalis-Menten Kinetic parameters
KM1=0.8;    % uM
KM2=0.1;
KM3=0.004;
vmax1=0.0018; % uM/s
vmax2=0.0018;
vmax3=0.000012;

%species equations
xdot = [-vmax1*x(1)/(x(1)+KM1) -
        vmax2*x(1)/(x(1)+KM2);
        vmax1*x(1)/(x(1)+KM1) - vmax3*x(2)/(x(2)+KM3);
        vmax3*x(2)/(x(2)+KM3)+vmax2*x(1)/(x(1)+KM2)];

```

The following figure shows a simulation of the changes in metabolite concentrations:

