# MATHEMATICAL MODELS OF GLUCOSINOLATE METABOLISM IN PLANTS

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf


Presented by

**Suraj Sharma**

from Sonbhadra, India

Düsseldorf, March 2018

Suraj Sharma: *Mathematical models of Glucosinolate metabolism in plants,*

SUPERVISORS:

Prof. Oliver Ebenhöh

Prof. Stanislav Kopriva

LOCATION:

Düsseldorf

## ABSTRACT

Glucosinolates are sulphur-rich secondary metabolites found in the plants of *Brassicacea* family. Glucosinolates play an important role in plants defence from pathogens. Depending on the type of microbes, specific glucosinolates can act as feeding deterrents or stimulants. A particular difficulty in the analysis of secondary metabolites is the vast diversity of different chemical structures. Considering the types of biochemical transformations, which could be involved in any secondary metabolite biosynthesis, in principle an infinite number of chemical structures could be produced. This is true for the experimental identification of secondary metabolites but the theoretical description of secondary metabolite biosynthesis is equally challenging. Developing models wherein all possible structures are represented as a unique variable is very challenging. In the present study, we developed a mathematical model of the chain-elongation pathway of aliphatic glucosinolates, which are derived from methionine, found in *Arabidopsis thaliana*. These glucosinolates show a structural diversity arising from the variation in the chain-elongation pathway taking place during biosynthesis. By providing the mathematical description of the rate laws governing the chain-elongation of aliphatic glucosinolates, we illustrate how the biosynthetic rates in the system depend on all other metabolite concentrations, a behaviour originating from the broad-range substrate specificity of the metabolic enzymes. Considering the pathway structure and the measured enzymatic properties, model simulation shows all characteristics of the actual differences between wild-type and mutants. The simulation allowed us to assess the individual effects of two processes—the knocking out of an enzyme and the compen-

satory expressions of other metabolic enzymes—that are difficult to dissect experimentally. The variation in glucosinolate concentration across Arabidopsis ecotypes could be a result of allelic compositions at different biosynthetic loci. By addressing the diversity induced by the chain-elongation process, our model illustrates how and why methionine-derived glucosinolates with a particular frequency are produced. Furthermore, by relating the allelic differences to the enzymatic properties, our model provides a theoretical framework to investigate how different metabolic phenotypes arise from genetic differences.

# ZUSAMMENFASSUNG

Glucosinolate sind schwefelreiche Sekundärmetabolite, die in den Pflanzen der Familie *Brassicacea* gefunden werden. Glucosinolate spielen eine wichtige Rolle bei der Abwehr von Pflanzen gegen Krankheitserreger. Abhängig von der Art der Mikroben, können bestimmte Glucosinolate als Abwehrsmittel oder Stimulanzien auf sie wirken. Eine besondere Schwierigkeit bei der Analyse von Sekundärmetaboliten ist deren große Vielfalt von vorhandenen chemischen Strukturen. Berücksichtigt man die Möglichkeiten der biochemischen Transformationen, die an der Biosynthese von Sekundärmetaboliten beteiligt sein könnten, wären im Prinzip eine unendliche Anzahl chemischer Strukturen erzeugbar. Dies gilt für die experimentelle Identifizierung von Sekundärmetaboliten, jedoch ist die theoretische Beschreibung der Biosynthese gleichermaßen anspruchsvoll. Modelle zu entwickeln, bei denen alle möglichen Strukturen als einzelne Variablen dargestellt werden, ist sogar unmöglich. In der vorliegenden Arbeit haben wir ein mathematisches Modell entwickelt, welches die Biosynthese von aliphatischen, Methionin abgeleiteten Glucosinolate in *Arabidopsis thaliana* beschreibt. Die strukturelle Vielfalt der Glucosinolate resultiert aus Variationen in der Kettenverlängerung während derer Biosynthese. Durch die mathematische Beschreibung der Geschwindigkeitsgesetze für die Kettenverlängerung von aliphatischen Glucosinolaten wird gezeigt, wie die Biosyntheseraten im System von allen anderen Metabolitkonzentrationen abhängen, ein Verhalten, das auf der breiten Substratspezifität der metabolischen Enzyme beruht. Unter Berücksichtigung der Syntheseweg-Struktur und der gemessenen enzymatischen Eigenschaften, zeigt

die Modellsimulation alle Merkmale der tatsächlichen Unterschiede zwischen Wildtyp und Mutanten. Die Simulation erlaubte es uns, die individuellen Auswirkungen von zwei Prozessen— dem Entfernen eines Enzyms und der kompensatorischen Expression anderer metabolischer Enzyme— zu untersuchen, die experimentell nur schwer zu analysieren sind. Die Variation in der Glucosinolatkonzentrationen in verschiedenen Arabidopsis-Ökotypen könnte das Ergebnis einer individueller Zusammensetzung von an der Biosynthese beteiligter Allele sein. Unter Berücksichtigung der durch den Kettenverlängerungsprozess induzierten Diversität veranschaulicht unser Modell, wie und warum Methionin abgeleitete Glucosinolate mit einer bestimmten Häufigkeit synthetisiert werden. Indem wir die allelischen Unterschiede auf die enzymatischen Eigenschaften beziehen, liefert unser Modell einen theoretischen Rahmen, um zu untersuchen, wie sich verschiedene metabolische Phänotypen aus genetischen Unterschieden ergeben.

# ACKNOWLEDGEMENTS

I am highly grateful to my supervisor Prof. Oliver Ebenhöh for giving me the opportunity to carry out my doctoral research under his guidance and supervision. I am very thankful to you for being such a great mentor and providing me with necessary facilities, which helped me in maintaining the smooth pace of my doctoral work.

I would like to express my keen gratification to my co-supervisor Prof. Stanislav Kopriva at the University of Cologne for his supervision and constant encouragement during the course of my Ph.D. work.

I am thankful to all the past and present members of Institute for Quantitative and Theoretical Biology for providing an excellent working environment. My special thanks to Dr Ovidiu Popa (*Alter*) for all the exciting conversations and suggestions during my Ph.D. work, and for caring like an elder brother. Next, I would like to thank Dr Anna Matuszyńska (*Captain*), Dr Antonella Succurro and Dr Adélaïde Raguin for having numerous healthy and fun-filled discussions. I would also like to thank Simon Schliesky and Jannina Mass for taking care of my IT infrastructure in due course of my Ph.D. journey. Also, I would like to thank Mara Schuff for taking care of all the bureaucratic paperwork during my Ph.D.

I would like to extend my gratification to all the members of Cluster of Excellence on Plant Sciences (CEPLAS) for their constant support and encouragement, and Deutsche Forschungsgemeinschaft (DFG) for the financial support. Also, I would like to thank Heinrich-Heine Universität, Düsseldorf for giving me the opportunity to pursue my doctoral degree.

I would like to express my warmest thanks to Aditi Sharma for the motivation, and for kindly reviewing a draft of the manuscript and giving many helpful suggestions.

Last but not the least, I would like to thank to my parents and all of my friends for their everlasting love, support and continuous encouragement. It would have not been possible without you all.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| AATR | Amino acid transporter |
| BAT5 | Bile acid transporter 5 |
| BCAT3 | Branched-chain aminotransferase 3 |
| BCAT4 | Branched-chain aminotransferase 4 |
| CYP79F1, CYP79F2 | Cytochrome P450 Dihomomethionine N-hydroxylase |
| GA | Genetic algorithm |
| Glc | $\beta - D - Glucopyranose$ |
| GSL | Glucosinolate |
| IPMI | Isopropylmalate isomerase |
| IPMDH | Isopropylmalate dehydrogenase |
| MAM1 | Methylthioalkylmalate synthase 1 |
| MAM2 | Methylthioalkylmalate synthase 2 |
| MAM3 | Methylthioalkylmalate synthase 3 |
| MDS | Multi-dimensional scaling |
| Met | Methionine |
| MTOB | 4-methylthio-2-oxobutanoate |
| MTOP | 5-methylthio-2-oxopentanoate |
| QTL | Quantitative trait locus |
| $A_i$ | amino acid $i$ |
| $A_{c,i}$ | cytosolic amino acid $i$ |
| $A_{p,i}$ | plastidic amino acid $i$ |
| $C_R$ | coefficient of response |
| $C^R_{p_k}$ | control coefficient for response R for parameter $p_k$ |

| | |
|---|---|
| $E_i$ | concentration of enzyme $i$ |
| $G_i$ | glucosinolate $i$ |
| $k_j^+, k_j^-$ | rate constants of reaction j in for the forward and backward direction |
| $K_{d_{i,j}}$ | dissociation constant with respect to substrate of chain-length j, catalysed by enzyme $i$ |
| $K_i$ | keto acid $i$ |
| $K_{c,i}$ | cytosolic keto acid $i$ |
| $K_{p,i}$ | plastidic keto acid $i$ |
| $M_{G,i}$ | metabolic genotype $i$ |
| $M_{P,i}$ | metabolic phenotype $i$ |
| $P_i$ | product $i$ |
| $S_i$ | substrate $i$ |
| $\nu$ | reaction rate |
| $\nu_{AATR,i}$ | rate of transport reaction $i$ mediated by AATR |
| $\nu_{BAT5,i}$ | rate of transport reaction $i$ mediated by BAT5 |
| $\nu_{BCAT3,i}$ | rate of reaction $i$ catalysed by BCAT3 enzyme |
| $\nu_{BCAT4,i}$ | rate of reaction $i$ catalysed by BCAT4 enzyme |
| $\nu_{CYP,i}$ | rate of reaction $i$ catalysed by CYP79 enzymes |
| $\nu_{MAM1,i}$ | rate of reaction $i$ catalysed by MAM1 enzyme |
| $\nu_{MAM3,i}$ | rate of reaction $i$ catalysed by MAM3 enzyme |
| $V_{max}$ | maximum velocity |
| $\Delta$ | euclidean distance |

# Part I

# GLUCOSINOLATES: IMPORTANCE AND BEYOND

Glucosinolates are a fascinating class of natural substances found in the plants of *Brassicaceae* family. Breakdown products of glucosinolates facilitate defence against plant pathogens. The distinct taste and flavour of certain *Brassica* vegetables (broccoli, cauliflower and cabbage) and condiments (mustard, horseradish and wasabi) is because of the presence of glucosinolates. For humans, glucosinolates functions as cancer-preventive agents and flavour compounds. Thus, to fully exploit the potential of glucosinolates in agriculture and medicine, complete understanding of why and how plants synthesise glucosinolates is important.

# 1

# INTRODUCTION

## 1.1 GLUCOSINOLATES: *what, why, where and how?*

Glucosinolates are sulphur-rich secondary metabolites, found in the plants of *Brassicaceae* family. Glucosinolates upon hydrolysis by endogenous thioglucosidases called myrosinases produce several different products (e.g., isothiocynates, thiocynates and nitriles) [36]. These products have different biological functions as defence compounds or attractants. Originally known as mustard oil glycosides, glucosinolates have been a part of human life for many centuries because of the strong flavours and tastes they elicit in *Brassica* vegetables that include cabbage, broccoli, and condiments like mustard and wasabi. Over the past few decades, the major focus of research have been on the negative aspects because of the prevalence of certain *antinutritional* or *goitrogenic* glucosinolates in the protein-rich meal from widely grown domesticated vegetable crops and condiments. However, there also exists a positive side, represented by the therapeutic and prophylactic properties of glucosinolates. The importance of glucosinolates has increased further following discovery of their potential as cancer-preventive agents, biofumigation, and crop-protection compounds [9, 31, 100].

*Glucoinolates: mustard oil glucosides*

### 1.1.1 *Chemical structure and hydrolysis*

Glucosinolates are anionic compounds, where a variable side chain (R) is attached to a common core glucosinolate structure that constitutes a β-D-Glucopyranose (Glc) residue linked to a (Z)-*N*-

hydroximininosulphate ester via a sulphur atom (see Figure 1). The

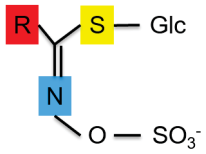



Figure 1: A schematic of the basic structure of a glucosinolate molecule. A variable side-chain (R) is attached to a core glucosinolate structure that constitutes a β-D-Glucopyranose (Glc) residue linked to a (Z)-*N*-hydroximininosulphate ester via a sulphur (S) atom.

first elucidated glucosinolate structures were of sinigrin and sinalbin in 1956 [20], however the term glucosinolate appeared first time in 1961 [19]. Till date, more than 135 different glucosinolates have been identified [2]. Glucosinolates are classified by their precursor amino acid. While glucosinolates derived from Alanine (Ala), Leucine (Leu), Isoleucine (Ile), Methionine (Met), or Valine (Val) are referred to as aliphatic, those derived from Phenyalanine (Phe) or Tyrosine (Tyr) are called aromatic, and those derived from Tryptophan (Trp) are called indolic glucosinolates. The variable group (-*R*) of most glucosinolates are elongated by one or more methylene moieties [22]. Both elongated and non-elongated *R* groups are subject to a wide range of modifications including hydroxylation, S-oxygenation, alkenylation, glycosylation, desaturation, and acylation [2, 22, 36].

Plants accumulating glucosinolates always possess an endogenous thioglucosidase known as myrosinase, which hydrolyses the glucose moiety on the main skeleton [83]. The hydrolysis products are glucose and an unstable aglycone that can rearrange to form products such as isothiocyanates, thiocyanates and nitriles (cf. Fig. 2). Formation of the hydrolysis products, however, depends on the physiological conditions under which the reaction takes place. The primary product at neutral pH is isothiocyanate, while under acidic conditions (pH $< 3$) and in the presence of ferrous ions or epithiospecifier proteins nitriles are formed [36, 61]. Hydrolysis in intact plants seems to be prevented by the spatial separation of glu-

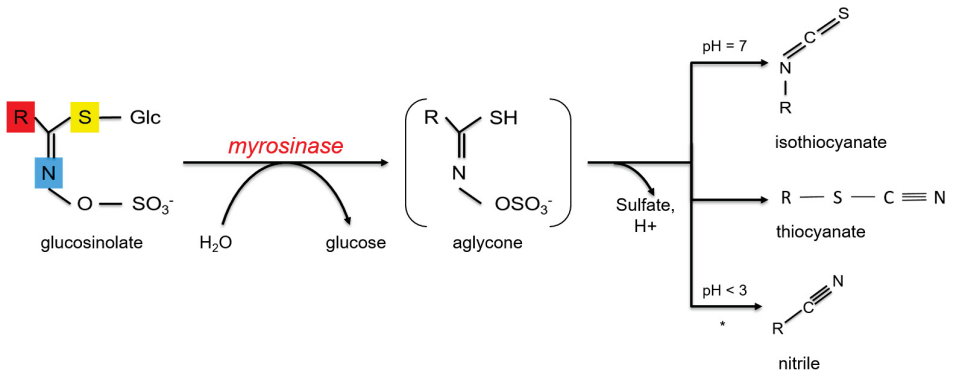



Figure 2: A schematic of the hydrolysis of a glucosinolate by myrosinase.

cosinolates and myrosinase. However, upon tissue damage these components mix together and lead to the rapid formation of glucosinolate hydrolysis products [36, 106]. The biological activities of glucosinolates are attributed to the activities of their hydrolysis products [106].

### 1.1.2 *Biological function of glucosinolates*

The activities of glucosinolate hydrolysis products upon plant damage have suggested that the major function of these glucosinolates in plants is to defend against herbivores and pathogens. Although the myrosinase system has been actively investigated as a feature of plant defence for about 100 years, there are still many gaps in our knowledge. The current research is replete with studies that demonstrate the outright toxicity, growth inhibition, or feeding deterrence induced by glucosinolates to a wide range of potential herbivores, including mammals, birds, insects, mollusks, aquatic invertebrates, nematodes, bacteria, and fungi (e.g., [8, 62, 101]). Based on current knowledge [8, 65, 98], isothiocyanates are known to be frequently responsible for the activity of the hydrolysed glucosinolates. However, little is known about the specific mechanism by which isothiocyanates exert their toxicity aside from their general propensity to react with amino and sulfhydryl groups of proteins in vitro [51].

Several studies demonstrate the toxicity of glucosinolate hydrolysis products to bacteria and fungi *in vitro* [6, 68, 94]. Thus, glucosinolates could be expected to defend plants against pathogens. Although resistance to pathogens *in vivo* is positively correlated with glucosinolate content, a defensive role cannot be assumed. It stems from the fact that many pathogens, especially biotrophic organisms, may not cause enough cell damage to activate the glucosinolate-myrosinase system.

Depending on the herbivore, the same glucosinolate that serve as general deterrents can also function as attractants. Many insect herbivores have specialised on glucosinolate-containing plants, and often use these compounds as cues for feeding or oviposition [28, 70, 86]. Herbivores that specialise on glucosinolate-containing plants must have some mechanism to overcome the toxicity induced by the glucosinolate hydrolysis products. Theoretically, they could instantly excrete glucosinolates from their bodies, metabolise them into non-toxic compounds, or be insensitive to the toxic actions of the glucosinolates. The larvae of two lepidopteran species are known to employ metabolic strategies to circumvent the toxicity of glucosinolates [84, 106]. Both species bypass the formation of isothiocyanate in their digestive tracts.

*Evidence for insect attraction comes from electrophysiological investigations, wherein the receptor organs (cells) respond directly to glucosinolates.*

Different glucosinolates seem to have varying effects on different herbivores [6, 87]. Once glucosinolates are released into the soil due to the decay of plant organs, or from root exudates, they may have important effects on the rhizosphere community. Studies have shown that the soil fungal species near glucosinolate-containing plants are different than the fungal species found elsewhere and exhibit increased tolerance to isothiocyanates [48].

### 1.1.3 *General biosynthesis of glucosinolates*

Biosynthesis of glucosinolates constitutes three independent steps: (i) chain elongation of selected precursor amino acids (only Met and Phe), (ii) formation of the core glucosinolate structure, and (iii) secondary modifications of the amino acid side chain. A schematic representation of the general biosynthesis of aliphatic glucosinolates, which are derived from Met, is shown in Figure 3. Together with the side-chain elongation, secondary modifications are responsible for more than 135 known glucosinolate structures [2], of which Arabidopsis has about 40, mainly derived from Met and Trp [52].

#### 1.1.3.1 *Chain elongation of Methionine*

Before entering the core structure pathway, methionine (Met) undergoes chain-elongation (Figure 3a). The process starts in the cytosol, with the deamination of Met by a branched-chain amino acid aminotransferase (BCAT) and thus forming a 2-oxo acid (Fig. 3 Step 1 to 2) [92]. The 2-oxo acid then enters a cycle of three successive transformations (Fig. 3 Steps 3 to 5): condensation with acetyl-CoA by a methylthioalkylmalate synthase (MAM), isomerisation by an isopropylmalate isomerase (IPMI), and oxidative decarboxylation by an isopropylmalate dehydrogenase (IPM-DH). The product of these steps is a 2-oxo acid that has been elongated by a methylene group (–CH2–). Hereupon, the molecule can either be transaminated by a plastid-localised BCAT to yield homo-methionine (homo-Met) (Fig. 3 Step 3 to 6) and enter the core glucosinolate structure pathway or proceed through another round of chain elongation. Consequently, the chain-elongation process yields various homo-Met of increasing chain-lengths.

The cloning studies of the GS-ELONG quantitative trait locus (QTL), which is known to control the variation in the side chain length of aliphatic glucosinolates, led to the identification of three

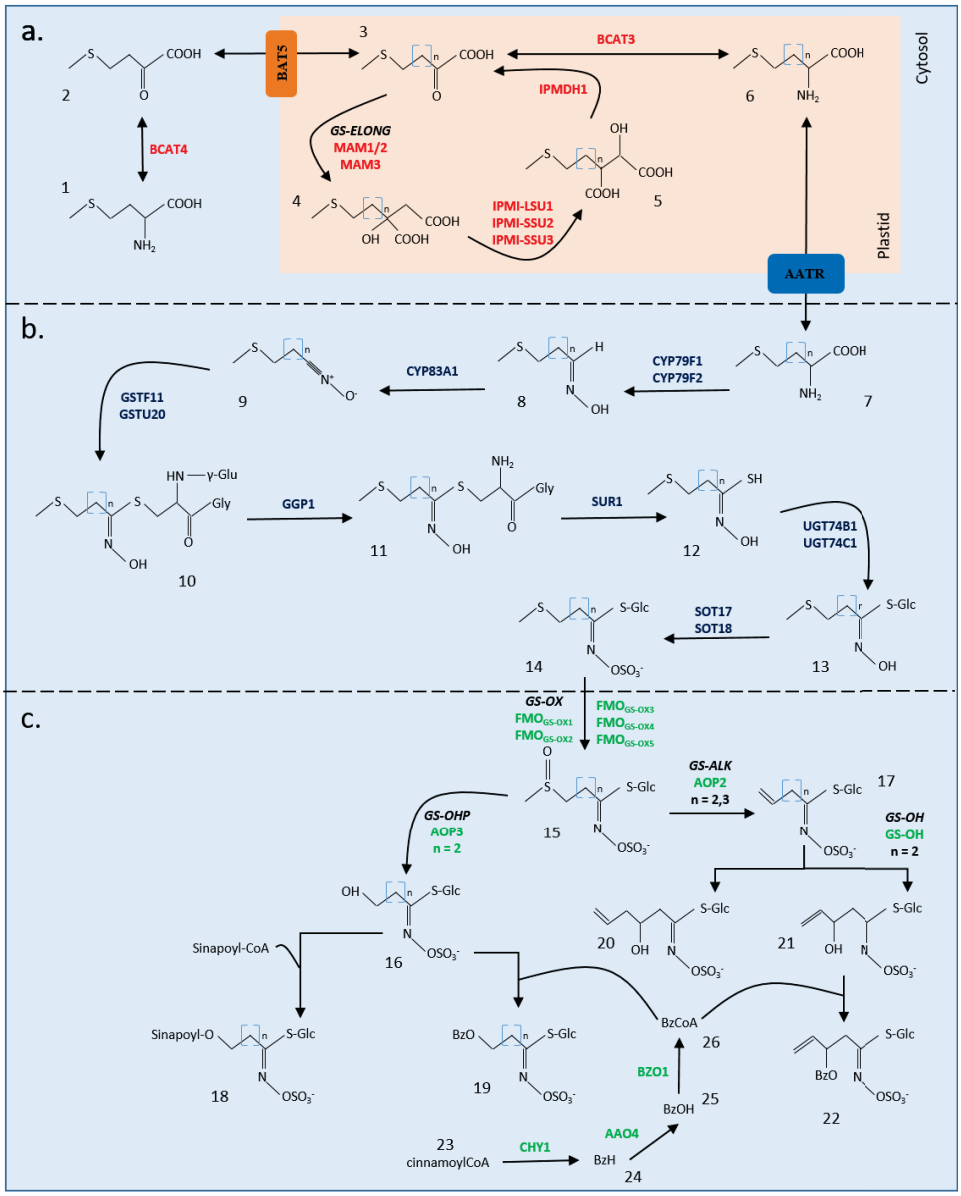



Figure 3: Schematic diagram of the chain-elongation of aliphatic glucosinolates, which are derived from methionine. (a) Chain-elongation of Met, (b) Constructing the GSL core, and (c) Secondary modifications. Metabolite 3 represents both the entry substrate and the chain-elongated product of each elongation cycle. The different steps are numbered and explained in detail in the main text. Enzymes are colour coded based on their activity in different steps of biosynthesis: red, chain elongation pathway; blue, core structure formation; and green, secondary modification. QTLs are marked in black italics.

genes MAM1, MAM2 and MAM3 [58, 59]. *In vitro*, MAM1 is able to catalyse the condensations in the first three elongation cycles, MAM2 only the first one, and MAM3 the first six [5, 96, 97]. Additionally, knockout analyses showed that MAM2 plays an important role in the production of aliphatic glucosinolates derived from a single elongation cycle, whereas MAM1 is involved in the production of glucosinolates derived from two elongation cycles, and MAM3 contributes to the production of all aliphatic glucosinolates [58, 59, 97]. IPMIs catalysing the isomerisation reaction (Fig. 3, step 4 to 5) are composed of a large and a small subunits. Knockdown mutants of the only gene encoding for the IPMI large subunit (IPMI-LSU1) accumulated both the putative substrate in Leu biosynthesis, 2-isopropylmalate and an oxidated derivative of the corresponding intermediate in the first chain elongation cycle of Met, 2-(3 o-methylsulfinyl)propylmalate [56, 89]. This suggests that IPMI-LSU1 plays a role in aliphatic glucosinolate biosynthesis. The IPMI small subunit (IPMI-SSU1) is encoded by three genes. Single knockout mutants of the two other genes, IPMI-SSU2 and IPMI-SSU3, revealed small changes in amino acid and glucosinolate levels in both leaves and seeds. However, their co-expression with glucosinolate biosynthetic genes implies a role in glucosinolate biosynthesis [56, 89]. IPMDH1 was identified based on the knowledge of Leu biosynthesis and strong co-expression with glucosinolate biosynthetic genes [40, 44, 89]. Moreover, an *ipmdh1* knockout mutant showed a decrease in glucosinolate content [40, 89].

#### 1.1.3.2 *Constructing the glucosinolate core*

The precursor amino acid derivatives are converted to aldoximes by cytochromes P450 of the CYP79 family (Fig. 3 Step 7 to 8). CYP79F1 converts all chain-elongated Met derivatives, and CYP79F2 only converts the long-chained Met derivatives [10, 39]. Further, the aldoximes are oxidised to activated compounds by cytochromes

P450 of the CYP83 family, where CYP83A1 converts aliphatic aldoximes [3, 39, 78] (Fig. 3 Step 8 to 9). Following conjugation of the activated aldoximes to a sulphur donor, the produced S-alkylthiohydroximates are converted to thiohydroximates by the C-S lyase SUR1 [73] (Fig. 3 Step 9 to 12). Thiohydroximates are in turn S-glucosylated by glucosyltransferases of the UGT74 family to form desulfoglucosinolates. UGT74B1 metabolise the Phe-derived thiohydroximates [17], and UGT74C1 has been suggested to glucosylate Met-derived substrates [29] (Fig. 3 Step 12 to 13). The glucosylation gives rise to desulfoglucosinolates, which are finally sulfated by the sulfotransferases SOT17 and SOT18 to form glucosinolates [81] (Fig. 3 Step 13 to 14). The sulphur donor that is conjugated to the activated aldoxime has long been thought to be Cys [33, 36]. However, recent investigations have made glutathione (GSH) a more likely sulphur donor [4, 91].

#### 1.1.3.3 *Secondary Modifications*

For aliphatic glucosinolates secondary modifications include oxygenations, hydroxylations, alkenylations and benzoylations (Figure 3c). QTL analyses have identified four gene loci (GS-ELONG, GS-OX, GS-AOP and GS-OH) as responsible for the side-chain variability of aliphatic glucosinolates in *Arabidopsis* [52]. Based on co-expression with aliphatic glucosinolate genes and the knowledge that FMO (flavin monooxygenase) catalyzes heteroatom oxygenation, the FMO GS-OX1, localised within the GS-OX locus, was identified as a candidate for S-oxygenation of aliphatic glucosinolates [38]. Moreover, phylogenetic analysis identified a Brassicaceae-specific subgroup of FMO genes, named FMO GS-OX1-5 [38, 64]. Enzymatic assays showed that FMO GS-OX1-5 S-oxygenate both short- and long-chain aliphatic glucosinolates, albeit with different chain-length specificity [38, 64] (Fig. 3 Step 14 to 15). GS-AOP is the collective QTL name of the linked loci GS-ALK and GS-OHP [52].

A fine-scale investigation identified two 2-oxoglutarate-dependent dioxygenases:AOP2 and AOP3. AOP2 catalyses the conversion of S-oxygenated glucosinolates to alkenyl glucosinolates (Fig. 3, step 15 to 17), whereas AOP3 catalyzes the conversion to hydroxyalkyl glucosinolates (Fig. 3, step 15 to 16) [52]. The GS-OH locus is responsible for the biosynthesis of the hydroxylated alkenyl glucosinolate 2-hydroxybut-3-enyl glucosinolate [76] (Fig. 3 metabolite 21). High levels of 2-hydroxybut-3-enyl glucosinolate pose one of the major obstacles for using Brassica crops as animal feed because its specific breakdown product, an oxazolidine-2-thione, causes goitre in pigs and poultry [24, 80]. Siliques and seeds of Arabidopsis contain benzoylated glucosinolates, whose formation requires the co-substrate benzoyl-CoA. Absence of benzoylated glucosinolates was observed in the knockout of CHY1, which seems to be responsible for the production of benzaldehyde based on *in vitro* enzyme activity [47]. In a screen for changes in seed glucosinolate content, benzoylated glucosinolates were absent from seeds of a mutant in the BZO1 gene [54]. Recombinant BZO1 could synthesize benzoyl-CoA from benzoic acid, implying BZO1 to function in production of benzoyl-CoA [54].

## 1.2 RESEARCH OBJECTIVES AND THESIS STRUCTURE

The aim of this study is to develop mathematical models of chain-elongation of aliphatic glucosinolates, found in *Arabidopsis thaliana*. This is an effort to investigate and understand the chain-length distribution of aliphatic glucosinolates and identify major regulators of flux through the biosynthetic pathway. The present study also aims to provide new insights of relating genetic variations to metabolic diversity, and to illustrate how exclusive patterns of glucosinolate accumulation emerge out of genetic variations.

Chapter 2 outlines the lineage of model construction traced from the fundamentals of kinetic modelling, and model assumptions. Starting from a general introduction to mathematical modelling and the chain-elongation pathway of Met-derived glucosinolates, it includes the mathematical description of rate laws governing the chain-elongation process. We extend the width of this chapter by presenting some model-based analyses to illustrate systemic properties of the chain-elongation pathway of Met-derived glucosinolates.

Chapter 3 deals with the application of developed model to reproduce patterns of glucosinolate concentrations in the leaves of different *Arabidopsis thaliana* ecotypes. Starting from simulating the glucosinolate profile of wild-type *Arabidopsis thaliana* Columbia ecotype, we further showcase some model-based analyses of the knockout mutants. By briefly introducing the concept of metabolic control analysis, we quantify the extent to which different metabolic enzymes limit the metabolic fluxes. This analysis serves to be highly relevant for experimental investigations, however, we discuss the theoretical aspects with application to the chain-elongation pathway of Met-derived glucosinolates in the wild-type *Columbia* ecotype.

Chapter 4 is meant to investigate the link between the metabolic genotypes and the associated phenotypes. By relating the allelic differences to metabolic diversity, we provide new insights of interpreting genomic variations in context of enzymatic properties of metabolic enzymes. By providing some results from bioinformatics analyses, we showcase how a metabolic phenotype is a consequence of an interplay of the genome architecture and expression of metabolic enzymes.

Chapter 5 summarises the key findings of this study and future directions. We discuss the importance and applications of mathematical modelling in pushing the envelope of our understanding of natural variation in the patterns of glucosinolate accumulation. We extend the width of the chapter by providing theoretical predic-

tions, which can be verified later by experiments, of how exclusive patterns of glucosinolate accumulation emerge out of genetic differences.

# Part II

# THE MODEL CONSTRUCTION

Any scientific description is a simplification, based on our understanding, of reality.

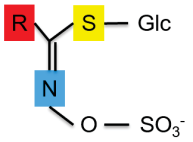
# 2

# MATHEMATICAL MODELLING OF THE CHAIN-ELONGATION OF ALIPHATIC GLUCOSINOLATES

Mathematical modelling is a powerful tool for the analysis of complex biological systems. It has been used extensively to study metabolic networks, resulting in a theory, termed metabolic control analysis, that describes in quantitative terms the role of metabolic enzymes in the regulation of pathway fluxes and metabolite concentrations [41]. Mathematical modelling is useful for describing experimental data, deducing regulatory principles, and understanding more complex dynamic phenomena such as oscillations in metabolic pathways (cf. [42]). Every mathematical model is based on some simplifying assumptions to facilitate the analytical or computational treatment and interpretation of the results. The iterative model-building process eliminates the unjustified assumptions, but a certain level of imprecision is deliberately accepted for simplicity.

Metabolic systems are defined by two types of data, where one encompasses the system variables like concentrations and fluxes, the other type comprises the system parameters such as stoichiometric coefficients and rate constants. The classical approach of metabolic modelling is concerned with simulating the time-dependent behaviour of the system variables for a given value of parameters by using ordinary differential equations. The approach dates back to the pioneering work on the glycolytic pathway [27]. The current research is replete with studies illustrating mathematical modelling of metabolic networks. A few examples of biochemical pathways are reminded in [11, 16, 42]. In this study, we use mathematical

modelling to study the chain-elongation of aliphatic glucosinolates, derived from methionine.




*Structure of a glucosinolate molecule*

Aliphatic glucosinolates are a major class of glucosinolates found in the model plant *Arabidopsis thaliana*. These are derived from methionine (Met) and have a variable side chain (R) attached to a common core glucosinolate structure, which constitutes a β-D-Glucopyranose (Glc) residue linked to a (Z)-*N*-hydroximininosulphate ester via a sulphur atom. The structural diversity of aliphatic glucosinolates arises from the variation in side-chain elongation and patterns of secondary oxidation and esterification [99]. The biosynthesis of Met-derived glucosinolates starts with the chain-elongation of Met, followed by the construction of core glucosinolate structure, and side chain modification [36]. As a major contributor to the diversity of glucosinolates, the chain-elongation pathway has gained considerable focus over the time. Met-derived glucosinolates in Arabidopsis have side-chain lengths of three to eight carbon atoms with the three- and four- carbon chain-lengths being predominant [45]. Studies have shown that chain elongation of Met involves a repetitive cycle of three reactions that result in the net addition of one methylene ($-CH_2-$) group for each cycle [12, 32, 63]. This cycle is known to run up to six times in *A. thaliana*, thus, yielding glucosinolates of six different chain-lengths [45].

In the present study, we focus on the kinetic modelling of the chain-elongation pathway of aliphatic glucosinolates, which are derived from methionine. Here, we describe the lineage of constructing the mathematical model of chain-elongation pathway traced from fundamental principles of deterministic kinetic modelling. Moreover, the model is used to investigate generic properties exhibited by the chain-elongation pathway. The results presented in this chapter are based on simulations using a reference set of kinetic parameters, wherein all of the metabolic enzymes are assumed to have equal catalytic efficiency, i.e. all dissociation constants, $K_d$ and maxi-

mum velocity, $V_{max}$ values corresponding to different substrates are set to 1.0 (arbitrary units). This is an effort to investigate and understand the systemic properties of the chain-elongation pathway. In this study, these parameters will be referred to as "vanilla" parameters.

## 2.1 THE FUNDAMENTALS OF KINETIC MODELLING

The principal notions are the concentration (i.e. number of moles of given substance per unit volume) and the reaction rates (expressed in terms of change in concentration per unit time). This type of modelling is also referred to as phenomenological modelling, as the molecules and their interactions are considered as fundamental concepts. Starting from general balance equations, the subsequent subsections highlight the fundamentals of kinetic modelling concerning rate laws and steady states of nonlinear enzymic systems.

### 2.1.1 *Balance Equations*

Chemical or biochemical kinetics are based on the postulate that the rate ($v$) of a reaction at a point $r = (x, y, z)$ in space at time $t$ can be expressed in terms of the concentration of all participating reactants.

$$v(r, t) = v[S(r, t), t], \tag{2.1}$$

where $S$ denotes the vector of concentration. An essential characteristics of metabolic networks is their stoichiometry. It indicates the molecularity with which substrates and products enter the reactions. The signs of the stoichiometric coefficients depend on the chosen orientation of the reaction. Production and consumption of metabolites by a reaction are denoted by positive and negative sto-

ichiometric coefficients, respectively. In case the (bio)chemical system constitutes more than one reaction, the reaction rates can be denoted by $\nu_j$ ($j = 1, \ldots, r$) and the stoichiometric coefficients by $n_{ij}$, where $i$ and $j$ refer to the subscripts of the substance and the reaction, respectively. In such cases, it is useful to arrange the stoichiometric coefficients in a matrix, where the rows refers to the substances, and the columns refer to the reactions.

When the (bio)chemical reactions are the only cause of change in concentration (i.e., there is no mass flow due to convection, diffusion, etc.), the temporal change in the concentrations is given by the balance equation

$$\frac{dS_i}{dt} = \sum_{j=1}^{r} n_{ij}\nu_j. \tag{2.2}$$

This equation is a consequence of conservation of mass, so that the contribution of all reactions can be summed. Equation 2.2 can be written in a matrix form as

$$\frac{dS}{dt} = N \cdot \nu, \tag{2.3}$$

where $S$, $N$, and $\nu$ denote the vector of concentrations, the stoichiometric matrix, and the vector of reaction rates, respectively.

### 2.1.2 *The Steady-state approximation*

The concept of steady state plays an important role in kinetic modelling. A metabolic system is said to subsist in a steady state if the metabolic variables (usually concentrations and fluxes) do not change within a tolerable accuracy over a certain time span of interest. As a matter of course, the concept of steady state is a mathematical idealisation that can describe real situations only in an approximative way, due to fluctuations of different nature. It implies that

the rate of formation of a metabolite is equal to the rate of its consumption. Therefore, at steady-state, the concentration of metabolite remains constant over time. Thus, equation 2.3 is represented as

$$N \cdot v = \frac{dS}{dt} = 0. \tag{2.4}$$

Equation 2.4 limits the solution space, however, many values of $v$ could still satisfy the equation. To restrict modelling analyses to essential features, one often investigates the potential solutions to the Eq. 2.4 satisfying some additional criteria.

Static situations are very common in biolgy. Popular examples include fairly constant body temperature of homeothermic animals, the glucosose concentration in blood and the pH in diverse range of living cells. Biochemical examples of virtually time-invariant states are reminded in [23, 42, 49, 105].

## 2.2 THE CHAIN-ELONGATION OF ALIPHATIC GLUCOSINOLATES

Based on current understanding [95], the biosynthesis of aliphatic glucosinolates, derived from methionine (Met), is highly compartmentalised in *Arabidopsis thaliana*. Figure 4 shows a schematic of the chain-elongation of Met-derived glucosinolates in *Arabidopsis thaliana*. It starts in the cytosol where the initial substrate Met is deaminated to form 4-methylthio-2-oxobutanoate (MTOB, a 2-oxo acid) by a branched-chain aminotransferase (BCAT4) [92]. MTOB, the initial substrate capable of undergoing elongation, is transported into the plastid by the bile acid transporter 5 (BAT5) [30]. In the plastid, methylthioalkylmalate synthases (MAM1, MAM2 and MAM3) catalyse a condensation reaction where MTOB reacts with acetyl-CoA to form 2-(2'methylthio)ethylmalate, which is isomerised to 3-(2'-methylthio)ethylmalate by isopropylmalate iso-

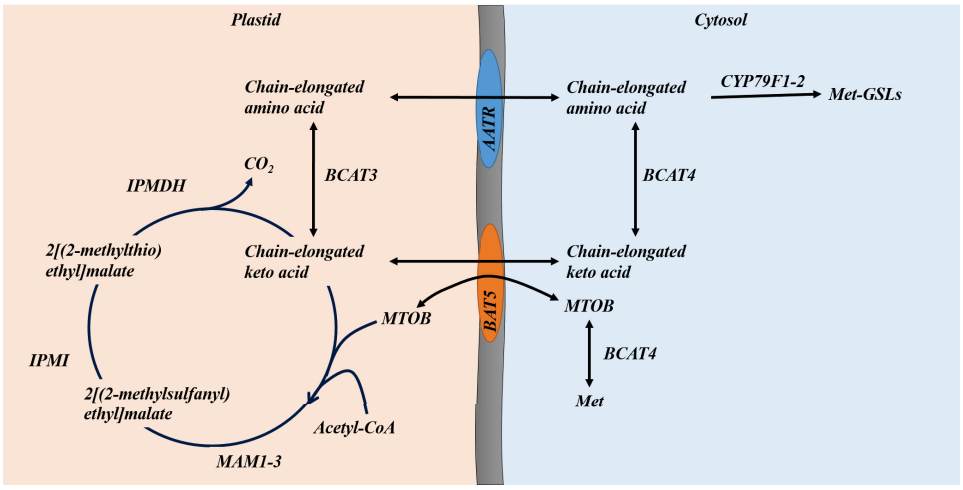



Figure 4: Schematic diagram of the chain-elongation of Met-derived glucosinolates. It starts with the deamination of Met by BCAT4 in cytosol. The product MTOB is transported into the plastid, where it goes through a three step cyclic chain-elongation process. The chain-elongated products can proceed through further steps of biosynthesis, and form Met-derived Met-derived glucosinolates. The different enzymatic steps are explained in detail in the main text.
AATR, amino acid transporter; *Met*, methionine; *MTOB*, 4-methylthio-2-oxobutanoate and; BAT5, bile acid transporter 5.

merase (IPMI) [12, 32, 63, 97]. The 3-alkylmalic acid is then decarboxylated by isopropylmalate isomerase dehydrogenase (IPMDH) to form 5-methylthio-2-oxopentanoate (MTOP) which is a homo-keto (2-oxo) acid, similar to MTOB, with an additional carbon in the side chain [32, 95, 97]. Hereupon, the chain-elongated keto acid has one of the following possibile fates: (i) proceed through a new cycle of chain elongation and thus forming a homo-keto acid of increasing side chain length, or (ii) get transaminated by a plastidic branched-chain aminotransferase (BCAT3) [55] and the resulting chain-elongated amino acid can proceed with the further steps of glucosinolate metabolism, or (iii) get exported by BAT5 to cytosol, where it is transaminated by BCAT4 or some other branched-chain aminotransferase to result in a chain-elongated amino acid, which can proceed to further steps of metabolism. For simplification, the transporter used for exporting the plastidic chain-elongated amino

acid to the cytosol is referred as amino acid transporter (AATR). The exact transporter is not yet identified.

## 2.3 THE MODEL SCHEME

For modelling purpose, we adopt the following simplifying assumptions: (i) the concentration of Met is maintained by a constant influx, (ii) the concentrations of amino-group donor and acceptor required for transamination reactions are assumed to be non-limiting and are not expressed explicitly in the model description, (iii) the Acetyl-CoA required for the condensation reaction, and the $CO_2$ produced during the decarboxylation reaction are assumed to have constant concentrations, and are not expressed explicitly in the model description, and lastly (iv) the series of three irreversible reactions of the chain-elongation cycle have been lumped to be represented as one. The rationale behind the last assumption arises from the fact that the enzyme catalysing the first irreversible reaction exerts full control over the metabolic flux in a branch, if there is no feedback from the downstream intermediates [41, 42]. Hence, all subsequent enzymes can be considered to exert virtually no control over flux, and the branch is effectively shortened.

In *Arabidopsis thaliana*, the chain elongation cycle is known to run for a maximum of six times, thus yielding six differentially-elongated products [45]. For mathematical convenience, the chain-elongation pathway of the Met-derived glucosinolates is redrawn as shown in Fig. 5. The influx of Met is denoted by $v_0$. The amino acid intermediates are denoted by $A_{c,i}$ and $A_{p,i}$, where $i$ corresponds to the number of additional carbon atoms, $c$ and $p$ denote the sub-cellular locations cytosol and plastid. Thus, $A_{c,0}$ denotes cytosolic methionine, $A_{c,1}$ cytosolic homomethionine, etc. Similarly, the keto acid intermediates are denoted by $K_{c,i}$ and $K_{p,i}$ with $K_{c,0}$ representing the cytosolic MTOB, $K_{c,1}$ cytosolic MTOP, etc. The three-step

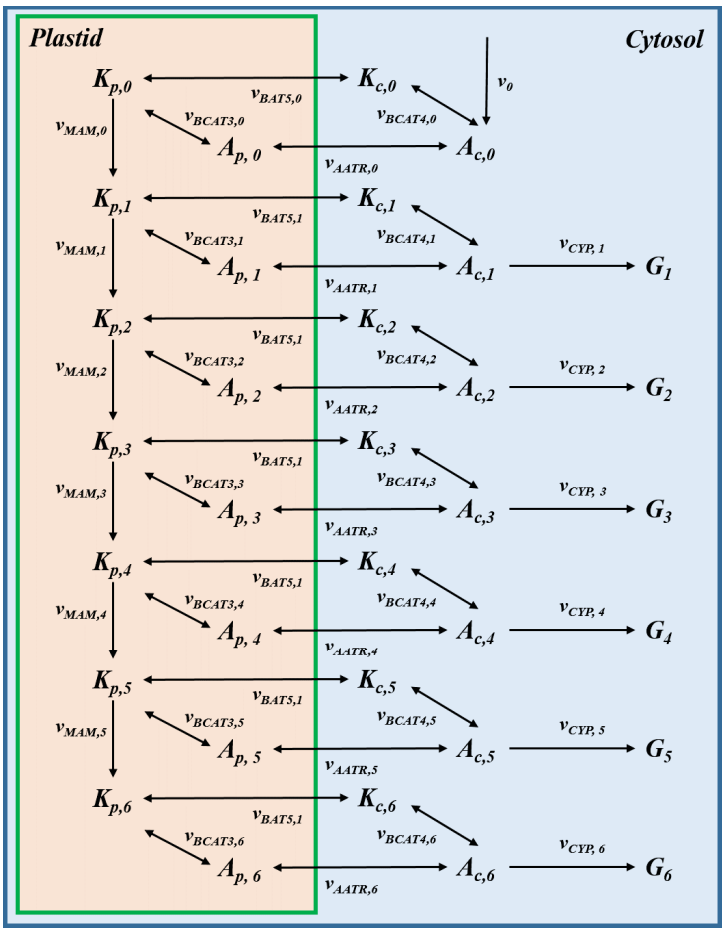



Figure 5: Schematic representation of the model of the chain-elongation of Met-derived glucosinolates. The influx of Met is denoted by $v_0$. The amino acid intermediates are denoted by $A_{c,i}$ and $A_{p,i}$. The subscripts c and p denote the sub-cellular location cytosol and plastid, respectively. Whereas i denotes the number of additional carbon atoms in its side-chain. Thus, $A_{c,0}$ denotes Met in cytosol. Similarly, keto acid intermediates in cytosol and plastid are denoted by $K_{c,i}$ and $K_{p,i}$, respectively, with $K_{c,0}$ representing MTOB. $G_i$ denotes the Met-derived glucosinolates. The three-step elongation cycle is represented by the reactions between $K_{p,i}$ and $K_{p,i+1}$. The downstream steps that produce glucosinolates are represented by the cytosolic reactions between $A_{c,i}$ and $G_i$, where $(1 \leqslant i \leqslant 6)$. Thus, $G_1$ denotes glucosinolate with one additional carbon atom in its side-chain, $G_6$ denotes the glucosinolate with six additional carbon atoms.

elongation cycle is represented by the reactions between $K_{p,i}$ and $K_{p,i+1}$. The rate of condensation reactions catalysed by MAM enzymes are denoted by $v_{K,i}$. Whereas, the rate of transamination reactions catalysed by the plastidic BCAT3 and the cytosolic BCAT4 enzymes are denoted by $v_{Ap,i}$ and $v_{Ac,i}$, respectively. The transport

reactions transporting keto acids $K_i$ and amino acids $A_i$ across the plastidial membrane are denoted by $v_{BAT5,i}$ and $v_{AATR,i}$, respectively. Lastly, the cytosolic chain-elongated amino acids $(A_{c,1}, \dots A_{c,6})$ are catalysed by CYP79 enzymes to form substrates that lead to the production of glucosinolates denoted by $G_i$. The rate of these reactions are denoted by $v_{G,i}$.

## 2.4 THE RATE LAWS

A very well-known and fundamental kinetic function is the mass-action rate law proposed by Guldberg and Waage (cf. [104]). It was derived from the idea that rate of a reaction is proportional to the probability of collision of reactants, which in turn is proportional to the concentration raised to the power of their molecularity, which is the number of molecules that have to meet to initiate the reaction.

However, the kinetics of an enzyme catalysed reaction exhibit features like saturation and inhibition that cannot, immediately, be described by mass-action kinetics. Saturation of enzymes, for example, arises from the fact that at high substrate concentration, nearly all enzyme molecules are bound to the substrate, so that a further increase in substrate concentration has almost no effect on reaction rate. To address such phenomena in enzyme kinetics Victor Henri, Leonor Michaelis and Maud Leonora Menten proposed a mechanism that was conceptually similar to mass-action law but was expressed in more precise mathematical and chemical terms, with an equilibrium between free enzyme and the enzyme-substrate and enzyme-product complexes [43, 72]. This mechanism is generally known as *Michaelis-Menten kinetics*. It applies to the enzymes that follow *one-enzyme-one-substrate* mechanism shown in scheme 2.5.

$$E + S \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} ES \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} E + P, \tag{2.5}$$

where $k_1^+$ and $k_2^+$ denote the rate of the reactions in forward direction, and $k_1^-$ and $k_2^-$ denote the reaction rates in reverse direction. While it was first derived for irreversible reactions, it was later generalised for reversible reactions [35].

Based on current understanding [55, 92], the glucosinolate biosynthetic enzymes generally possess a broad-range substrate specificity, which means an enzyme E exerts activity on different substrates denoted by $S_i (i = 1 \dots n)$ (cf. Scheme 2.6).

$$\begin{aligned}
E + S_1 &\underset{k_{1,1}^-}{\overset{k_{1,1}^+}{\rightleftharpoons}} ES_1 \underset{k_{1,2}^-}{\overset{k_{1,2}^+}{\rightleftharpoons}} E + P_1 \\
E + S_2 &\underset{k_{2,1}^-}{\overset{k_{2,1}^+}{\rightleftharpoons}} ES_2 \underset{k_{2,2}^-}{\overset{k_{2,2}^+}{\rightleftharpoons}} E + P_2 \\
&\vdots \\
E + S_n &\underset{k_{n,1}^-}{\overset{k_{n,1}^+}{\rightleftharpoons}} ES_6 \underset{k_{n,2}^-}{\overset{k_{n,2}^+}{\rightleftharpoons}} E + P_n
\end{aligned} \tag{2.6}$$

The time-dependent changes in the concentration of the enzyme-substrate complex are determined by

$$\begin{aligned}
\frac{dES_1}{dt} &= k_{1,1}^+ S_1 \cdot E - (k_{1,1}^- + k_{1,2}^+) ES_1 + k_{1,2}^- \cdot E \cdot P_1 \\
\frac{dES_2}{dt} &= k_{2,1}^+ S_2 \cdot E - (k_{2,1}^- + k_{2,2}^+) ES_2 + k_{2,2}^- \cdot E \cdot P_1 \\
&\vdots \\
\frac{dES_n}{dt} &= k_{n,1}^+ S_n \cdot E - (k_{n,1}^- + k_{n,2}^+) ES_n + k_{n,2}^- \cdot E \cdot P_n
\end{aligned} \tag{2.7}$$

Using the quasi steady-state assumption [72], one obtains the generalised rate law for the production of $P_i$ with $n$ competing substrates for the binding of enzyme E. A full derivation of the kinetic rate law for monomolecular reactions is given in Appendix B. The transamination reactions catalysed by BCAT3 and BCAT4, and the plastidial

transport mediated by BAT5 and AATR transporters are modelled by the reversible rate law, which has the general form [69, 90] as

$$v_{rev_i} = \frac{V^+_{m_i}\frac{S_i}{K^+_{d_i}} - V^-_{m_i}\frac{P_i}{K^-_{d_i}}}{1 + \sum_{j=1}^{n}\frac{S_j}{K^+_{d_j}} + \sum_{j=1}^{n}\frac{P_j}{K^-_{d_j}}}, \tag{2.8}$$

where $V^+_{m_i}$ and $V^-_{m_i}$ denote the maximum velocities of the reactions catalysing substrate $S_i$ and product $P_i$, respectively. Similarly, $K^+_{d_i}$ and $K^-_{d_i}$ are the dissociation constants of the reactions catalysing substrate $S_i$ and product $P_i$, respectively. Whereas $K^+_{d_j}$ and $K^-_{d_j}$ describe the dissociation constants with respect to substrates $S_j (j = 1 \ldots n)$ and products $P_j (j = 1 \ldots n)$, respectively.

The condensation and outflux reactions catalysed by MAM and CYP79 enzymes, respectively, are modelled by the irreversible rate law, which has the general form [13] as

$$v_{irr_i} = \frac{V_{m_i}\frac{S_i}{K_{d_i}}}{1 + \sum_{j=1}^{n}\frac{S_j}{K_{d_j}}}, \tag{2.9}$$

where $V_{m_i}$ and $K_{d_i}$ are the maximum velocity and dissociation constant, respectively, of the reaction catalysing substrates $S_i$. Whereas $K_{d_j}$ describe the dissociation constants of the respective substrates $S_j (j = 1 \ldots n)$.

The rate laws, which are used to construct the mathematical model of chain-elongation of Met-derived glucosinolates are given in Table 1.

| REACTIONS | RATE LAWS |
|---|---|
| $K_{p,i} \xrightarrow{MAM_1} K_{p,i+1}$ | $v_{K,i} = \frac{V_{m_i} \frac{Kp_i}{K_{d_i}}}{1 + \sum_{j=0}^{1} \frac{Kp_j}{K_{d_j}}}$ |
| $K_{p,i} \xrightarrow{MAM_3} K_{p,i+1}$ | $v_{K,i} = \frac{V_{m_i} \frac{Kp_i}{K_{d_i}}}{1 + \sum_{j=0}^{5} \frac{Kp_j}{K_{d_j}}}$ |
| $A_{c,i} \xrightleftharpoons{BCAT_4} K_{c,i}$ | $v_{Ac,i} = \frac{V_{m_i^+} \frac{Ac_i}{K_{d_i^+}} - V_{m_i^-} \frac{Kc_i}{K_{d_i^-}}}{1 + \sum_{j=0}^{6} \frac{Ac_j}{K_{d_j^+}} + \sum_{j=0}^{6} \frac{Kc_j}{K_{d_j^-}}}$ |
| $A_{p,i} \xrightleftharpoons{BCAT_3} K_{p,i}$ | $v_{Ap,i} = \frac{V_{m_i^+} \frac{Ap_i}{K_{d_i^+}} - V_{m_i^-} \frac{P_i}{K_{d_i^-}}}{1 + \sum_{j=0}^{6} \frac{Ap_j}{K_{d_j^+}} + \sum_{j=0}^{6} \frac{Kp_j}{K_{d_j^-}}}$ |
| $A_{c,i} \xrightleftharpoons{AATR} A_{p,i}$ | $v_{AATR,i} = \frac{V_{m_i^+} \frac{Ac_i}{K_{d_i^+}} - V_{m_i^-} \frac{P_i}{K_{d_i^-}}}{1 + \sum_{j=0}^{6} \frac{Ac_j}{K_{d_j^+}} + \sum_{j=0}^{6} \frac{Ap_j}{K_{d_j^-}}}$ |
| $K_{c,i} \xrightleftharpoons{BAT_5} K_{p,i}$ | $v_{BAT5,i} = \frac{V_{m_i^+} \frac{Kc_i}{K_{d_i^+}} - V_{m_i^-} \frac{P_i}{K_{d_i^-}}}{1 + \sum_{j=0}^{6} \frac{Kc_j}{K_{d_j^+}} + \sum_{j=0}^{6} \frac{Kp_j}{K_{d_j^-}}}$ |
| $A_{c,i} \xrightarrow{CYP_{79}F_1} G_i$ | $v_{G,i} = \frac{V_{m_i} \frac{Ac_i}{K_{d_i}}}{1 + \sum_{j=1}^{6} \frac{Ac_j}{K_{d_j}}}$ |
| $A_{c,i} \xrightarrow{CYP_{79}F_2} G_i$ | $v_{G,i} = \frac{V_{m_i} \frac{Ac_i}{K_{d_i}}}{1 + \sum_{j=5}^{6} \frac{Ac_j}{K_{d_j}}}$ |

Table 1: Rate laws used in building the chain-elongation model. The cytosolic amino acid intermediate *i* is denoted by $A_{c,i}$, while the plastidic amino acid intermediate is denoted by $A_{p,i}$. Similarly, $K_{c,i}$ and $K_{p,i}$ denote the cytosolic and plastidic keto acid intermediate *i*, respectively.

## 2.5 RESULTS

### 2.5.1 *Model formulation*

Based on current understanding [45], the maximum number of elongation cycle in *Arabidopsis thaliana* is six. Moreover, there exist isoforms for some enzymes [95], for example, there exist three MAM synthases, *MAM1*, *MAM2* and *MAM3*. While MAM3 catalyses the condensation reaction of keto acid intermediates of different chain lengths ($K_{p,0} \dots K_{p,5}$), the isoforms MAM1 and MAM2 catalyse the condensation of MTOB ($K_{p,0}$) and MTOP ($K_{p,1}$), only [58, 59, 97]. However, the majority of Arabidopsis ecotypes possess either MAM1 or MAM2 [59]. There also exist two CYP79 enzymes, CYP79F1 and CYP79F2, where the former accepts as substrates all chain-elongated amino acids ($A_{c,1} \dots A_{c,6}$) and the latter only pentahomomethionine ($A_{c,5}$) and hexahomomethioinine ($A_{c,6}$) [10]. Also, we consider two isoforms of brached-chain aminotransferases BCAT3 and BCAT4 active in the plastid and cytosol, respectively. We assume that both, BCAT4 and BCAT3, can deaminate amino acids of variable chain lengths. Lastly, we also consider the activity of two transporters BAT5 and AATR, where the former mediates the transport of keto acid intermediates of variable chain lengths, the latter mediates the transport of amino acid intermediates of all chain lengths across the plastidial membrane. Thus, our model constitutes the activity of six enzymes, namely MAM1, MAM3, BCAT3, BACT4, CYP79F1, and CYP79F2, and two transporters, namely BAT5, and AATR. These have a broad-range substrate specificity and exert activity on 28 intermediates ($A_0 \dots A_6$ and $K_0 \dots K_6$). These 28 metabolites act as substrate in 72 reactions (see Table 3 for details) taking place across two sub-cellular compartments, i.e. plastid and cytosol.

The mathematical description of the rate laws is derived from Michaelis-Menten kinetics to reflect the broad-range substrate speci-

ficity of metabolic enzymes. The rate of $i^{th}$ cytosolic and plastidic transamination reaction is denoted by $v_{BCAT4,i}$ and $v_{BCAT3,i}$, respectively. The MAM-catalysed $i^{th}$ condensation reaction rate is denoted by $v_{MAM,i}$, while the CYP79-catalysed $i^{th}$ reaction rate is given by $v_{CYP,i}$. Lastly, $v_{BAT5,i}$ and $v_{AATR,i}$ denote the $i^{th}$ transport reactions with keto and amino acid intermediates, respectively. Refer Table 1 for details.

The time-dependent changes in the concentrations of $i^{th}$ amino acid intermediate ($A_{c,i}$, $A_{p,i}$) and keto acid intermediate ($K_{c,i}$, $K_{p,i}$) are calculated by the following ordinary differential equations (ODEs):

$$\frac{dA_{c,i}}{dt} = -v_{BCAT4,i} - v_{AATR,i} - v_{CYP,i} \quad (\text{for} \quad 1 \leqslant i \leqslant 6) \tag{2.10}$$

$$\frac{dK_{c,i}}{dt} = v_{BCAT4,i} - v_{BAT5,i} \qquad (\text{for} \quad i = 0, \ldots, 6) \tag{2.11}$$

$$\frac{dA_{p,i}}{dt} = v_{AATR,i} - v_{BCAT3,i} \qquad (\text{for} \quad i = 0, \ldots, 6) \tag{2.12}$$

$$\frac{dK_{p,i}}{dt} = v_{BAT5,i} + v_{MAM,i-1} - v_{MAM,i} - v_{BCAT3,i} \quad (\text{for} \quad 1 \leqslant i \leqslant 5). \tag{2.13}$$

Whereas, for $i = 0$, the temporal change in the concentrations of $A_{c,i}$ and $K_{p,i}$ is given by

$$\frac{dA_{c,i}}{dt} = v_0 - v_{BCAT4,i} - v_{AATR,i} \tag{2.14}$$

$$\frac{dK_{p,i}}{dt} = v_{BAT5,i} - v_{MAM,i} - v_{BCAT3,i} \tag{2.15}$$

However, the temporal change in the concentrations of $K_{p,i}$ for $i = 6$ is calculated by

$$\frac{dK_{p,i}}{dt} = v_{BAT5,i} + v_{MAM,i-1} - v_{BCAT3,i}. \tag{2.16}$$

At steady-state, equations 2.10, 2.11, 2.12, 2.13, 2.14, 2.15 and 2.16 are reduced to the following equations:

$$\begin{aligned} \frac{dA_{c,i}}{dt} &= 0 \\ \frac{dK_{c,i}}{dt} &= 0 \\ \frac{dA_{p,i}}{dt} &= 0 \\ \frac{dK_{p,i}}{dt} &= 0. \end{aligned} \tag{2.17}$$

### 2.5.2 *Effect of varying influx*

Influx of metabolites plays an important role in governing the steady-state concentrations in an enzymatic system [41]. Thus, to study the effect of varying influx, we used *vanilla* parameters (Table 4) to simulate the steady-states for different values of Met influx ($v_0$). Figure 6 shows the steady-state concentrations of keto acid intermediates $K_{c,i}(i = 1, \ldots, 6)$, simulated for different Met influx $v_0$ values. With increasing influx concentration, we see a nonlinear increase in the steady-state concentrations of the keto acid intermediates. Around 15 fold higher concentration, in comparison to $v_0$, of intermediates is observed at $v_0 = 0.45$. However, a further increase in $v_0$ does not yield a steady-state. This behaviour is often known as substrate overloading, wherein the metabolic enzymes are saturated resulting in high accumulation of substrates.

Conversely, when we studied the effect of varying influx $v_0$ on the steady-state outfluxes ($v_{cyt,i}$ in Fig. 5) that contributes to the production of different Met-derived GSLs ($3C, \ldots, 8C$), we observed an in-

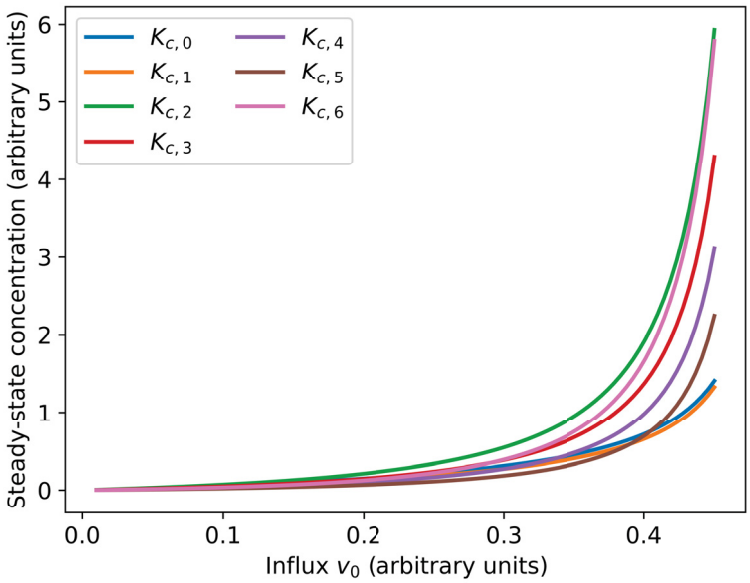



Figure 6: Effects of varying influx on the steady-state concentrations of keto acid intermediates. While *x-axis* quantifies influx $v_0$ concentration, *y-axis* quantifies the steady-state concentration of $K_{c,i} (i = 0, \ldots, 6)$. The subscript c stands for cytosol, whereas the numbers stand for additional number of carbon atoms in the side-chain of the keto acid intermediate (cf. Fig. 5). Values are given in arbitrary units. Results are based on the simulations performed using *vanilla* parameters (Table 4).

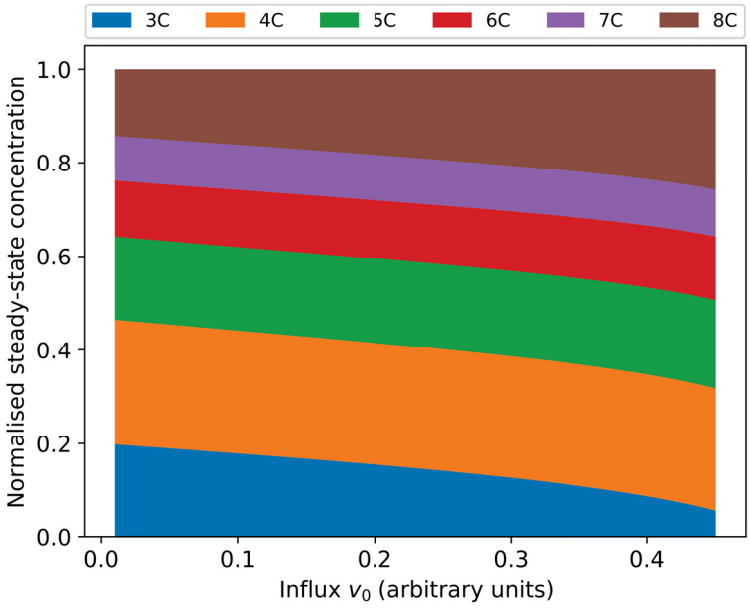



Figure 7: Effects of varying influx $v_0$ on the normalised steady-state concentrations of Met-derived glucosinolates (3C, ..., 8C). The values on *y-axis* are normalised to $v_0$ plotted on *x-axis*. Results are based on the simulations performed using *vanilla* parameters (Table 4).

teresting phenomenon. The steady-state flux can be compared with the steady-state concentrations if we assume that the steady-state flux is diluted by plants growth. Figure 7 shows the normalised steady-state concentrations of Met-derived GSLs at different concentrations of $v_0$. Fig. 7 clearly shows that with increasing influx $v_0$ concentration, the chain-length distribution shifts toward higher production of long-chain (8C) GSLs. Whereas, a reduction in 3C GSLs is also observed.

### 2.5.3 *Effect of transporters*

To study the behaviour originating from an extra cellular compartment, we consider the following two scenarios: (i) the reactions of chain-elongation of Met-derived GSLs take place in respective compartments (cf. Fig. 5), and (ii) all the reactions of the chain-elongation pathway take place in one compartment. The two models depicting the two aforementioned scenarios will referred to as "Generic" and "Compartmentalised" models.

Figure 8 shows a schematic of the generic model, where all the reactions are assumed to take place in one compartment. In such a case, the chain-elongation pathway can be represented as an interplay of 14 metabolic intermediates ($A_{0,\ldots,6}$ and $K_{0,\ldots,6}$). While A and K, respectively, denote the amino and keto acid intermediates, G represents the Met-derived glucosinolates. The numbers in subscript denote the additional number of carbon atoms in the molecule.

We simulated the steady-states at $v_0 = 0.4$ using *vanilla* parameters (Table 4) for the set of reactions depicted in two model schemes shown in Fig. 5 and 8. The steady-state outflux from the *generic* and *compartmentalised* models will be referred to as $J_g$ and $J_c$, respectively, which comprise six flux values denoted by respective $v_{CYP,i} (i = 1, \ldots, 6$, cf. Fig. 5 and 8). Further, to investigate the role of transporters in regulating the steady-state fluxes, we simulated

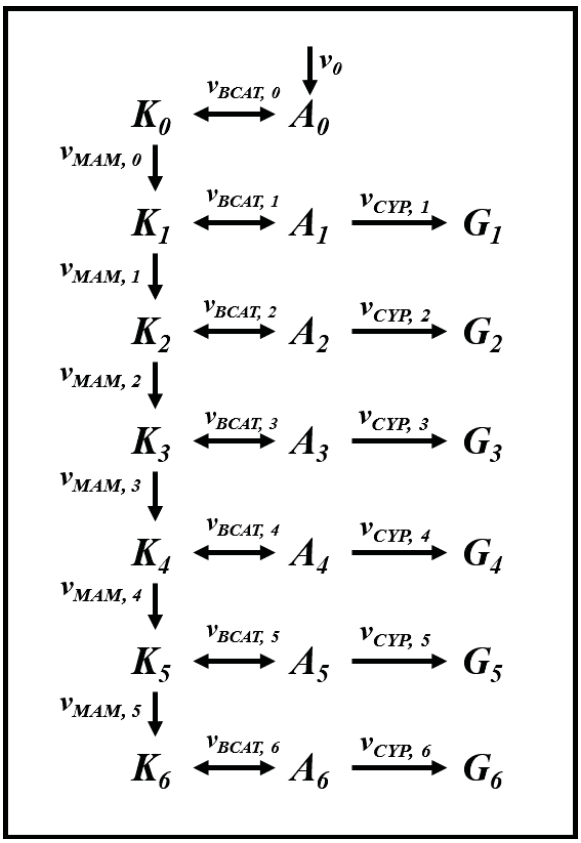



Figure 8: Schematic representation of the generic model of the chain-elongation of Met-derived glucosinolates. The chain-elongation pathway can be represented as an interplay of 14 metabolic intermediates ($A_{0...6}$ and $K_{0...6}$) if all the reactions of chain-elongation pathway takes place in one compartment. While A and K denote the amino and keto acid intermediates, respectively, G represents the Met-derived glucosinolates. The numbers in subscript denote the additional number of carbon atoms in the molecule.

the steady-states for different rate constants of the transport reactions, $v_{AATR,i}$ and $v_{BAT5,i}$, in the *compartmentalised* model. We varied the rate constants by varying the enzyme concentration ($E_t$), which changes the maximum velocity as

$$V_{max} = E_t \cdot k_{cat}, \tag{2.18}$$

where $k_{cat}$ is the rate by which the substrate is catalysed to form product.

Figure 9 shows a heatmap of the difference between the steady-state fluxes, $J_g$ and $J_c$, at different rate constants of the transport

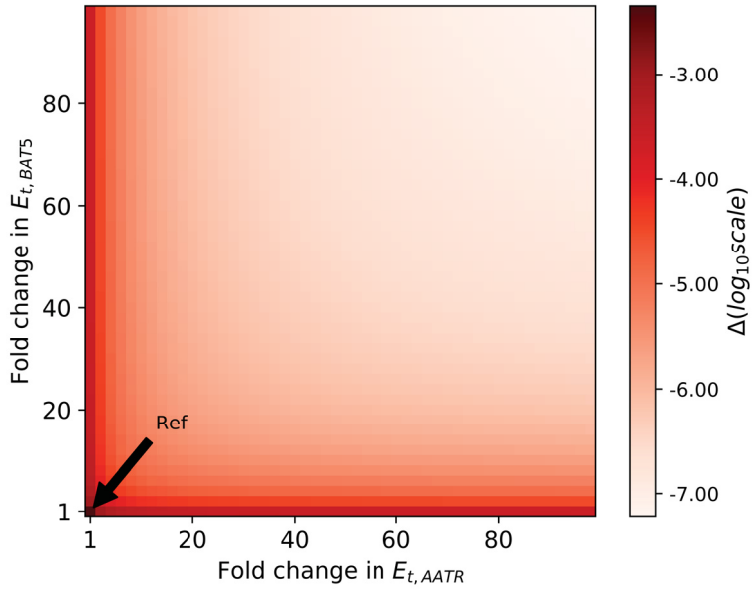
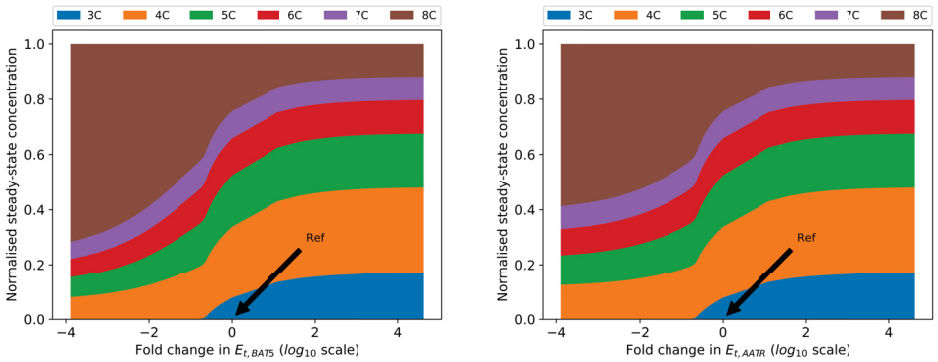



Figure 9: Heatmap of the difference between the steady-state fluxes from *generic* and *compartmentalised* model simulated for different rate constants $E_t$. The difference is measured in terms of euclidean distance $\Delta$, which is given in $log_{10}$ scale. Along the *x-axis* the concentration ($E_{t,AATR}$) of AATR is varied, whereas on *y-axis* the concentration ($E_{t,BAT5}$) of BAT5 transporter is varied. The values are presented as fold changes in the concentration of transporters at reference (Ref) state. Results are based on simulations performed using *vanilla* parameters (Table 4) for $v_0 = 0.4$ (arbitrary units).

reactions $v_{AATR,i}$ and $v_{BAT5,i}$. The difference between steady-state fluxes is measured as euclidean distance, $\Delta = \| J_g - J_c \|$. A decrease in $\Delta$ with increasing values of the $E_{t,BAT5}$ and $E_{t,AATR}$ is seen in Fig. 9. It highlights the fact that for cases where the transport reactions are very fast, the metabolites attain rapid equilibrium, and the metabolite concentrations in the cytosol is in equilibrium with plastidial concentrations. In such cases, it is plausible to assume the non-limiting effects of compartmentalisation.

A fine-scale investigation of the $J_c$ at different $E_{t,BAT5}$ and $E_{t,AATR}$ is shown in Fig. 10. The steady-state flux is comparable to the steady-state concentration if we assume that the steady-state flux is diluted by growth. To investigate the chain-length distribution, the steady-state values are normalised to the $v_0$ concentrations. At high $E_{t,BAT5}$ the steady-state chain-length distribution of GSLs shifts towards the




(a) (b)

Figure 10: Effects of varying concentrations of BAT5 and AATR transporters on the steady-state concentrations of different Met-derived GSLs (3C . . . 8C). Figure (a) shows the normalised steady-state concentrations of Met-derived GSLs at different concentrations ($E_{t,BAT5}$) of BAT5 transporter. Whereas, figure (b) shows the normalised steady-state concentrations of Met-derived GSLs at varying concentration ($E_{t,AATR}$) of AATR transporter.
The values on *x-axis* are presented as fold changes in the concentration of BAT5 and AATR transporters with respect to the reference concentration marked as "Ref". Results are based on simulations performed using *vanilla* parameters (Table 4).

production of short-chains GSLs (3C and 4C), while low $E_{t,BAT5}$ diverts the chain-length distribution towards the production of long-chain GSLs (refer Fig. 10a). A similar behaviour is exhibited, when we varied $E_{t,AATR}$, and is shown in Fig. 10b. A quantitative difference, however, in the chain-length distribution at low $E_{t,BAT5}$ and $E_{t,AATR}$ is observed.

### 2.5.4 *Effect of enzyme concentration*

Variation in glucosinolate accumulation at different developmental stages of *Arabidopsis thaliana* is often observed [7]. It could be a result of varying expression of metabolic genes, which can amount to the total concentration of respective enzymes. Thus, to study the effects of varying enzyme concentrations on the steady-state distribution of Met-derived GSLs of different chain-length, we used *vanilla* parameters to simulate the steady-state for different concentrations of

metabolic enzymes. According to Eq. 2.18, enzyme concentration ($E_t$) changes $V_{max}$ of respective enzyme-catalysed reactions.

Figure 11 shows the normalised steady-state concentrations of different Met-derived GSLs (3C, . . . , 8C) at different concentrations of metabolic enzymes: (a) MAM1, (b) MAM3, (c) BCAT3, (d) BCAT4, (e) CYP79F1, and (f) CYP79F2. To reflect the change in concentration, the values on *x-axis* is given as fold changes in the concentration to the reference marked with "Ref". Figure 11a shows that at low concentration of MAM1 chain-length distribution shifts toward high production of 3C GSLs. Whereas high concentration of MAM1 shifts the distribution towards higher production of 4C GSLs than 3C. Next, in Fig. 11b we show how variation in the concentration of MAM3 regulates the steady-state chain-length distribution. We see that at low MAM3 concentration the chainlength distribution is favoured towards the production of short-chain GSLs, while high concentration favour the production of long-chain GSLs. Conversely, Fig. 11c and 11d show that a low concentrations of transaminases, BCAT3 and BCAT4, shifts the steady-state chain-length distribution toward high concentrations of long-chain GSLs. Whereas high concentrations of transaminases favour high production of short-chain GSLs. Lastly, the effect of varying concentration of CYP79 enzymes is recorded in Fig. 11e and 11f. While low concentration of CYP79F1 favours the high production of 7C and 8C GSLs, high concentration of CYP79F1 enzyme leads to a shift towards the short-chain GSLs (cf. Fig. 11e). Whereas, varying the concentration of CYP79F2 enzymes shows almost no effect on the chain-length distribution of GSLs.

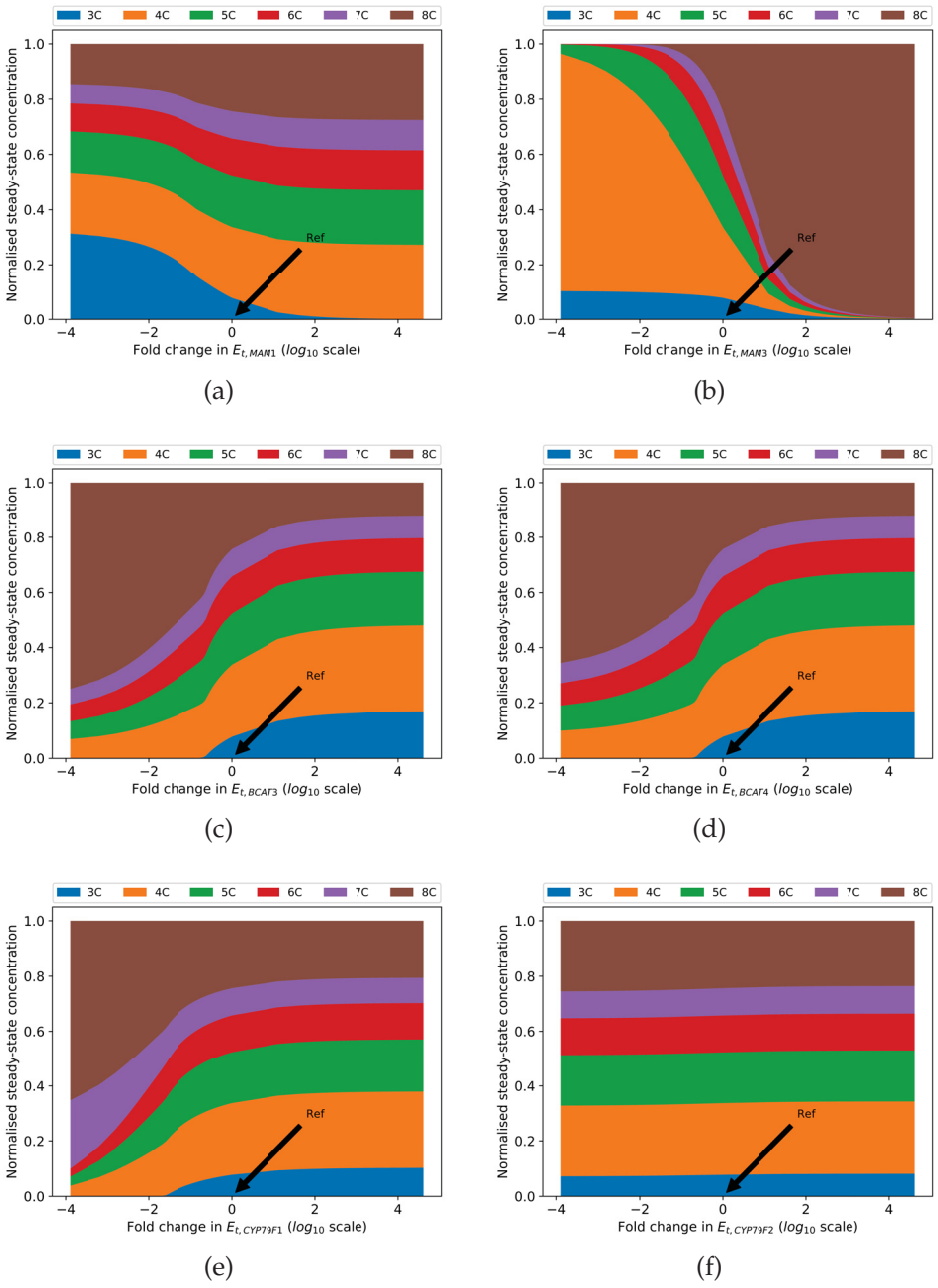



Figure 11: Effect of varying enzyme concentrations ($E_t$) on the steady-state concentrations of Met-derived glucosinolates (3C, . . . , 8C). Normalised steady-state concentration of different GSLs at different concentrations of (a) MAM1, (b) MAM3, (c) BCAT3, (d) BCAT4, (e) CYP79F1, and (f) CYP79F2 is presented. The values on *y-axis* is normalised to the influx concentration. Whereas, the values on *x-axis* is presented as fold changes in $E_t$ with resepct to reference marked as "Ref". Simulations are performed using *vanilla* parameters (Table 4).

### 2.5.5 *Effect of enzyme specificity*

In 1995, Radzicka and Wolfenden [82] introduced the idea of catalytic efficiency, as a measure of the capacity of an enzyme to accelerate a reaction beyond its uncatalysed rate. Specificity is a fundamental property of an enzyme that does not illustrate how enzymes are excellent catalysts for certain reactions but how they are extremely poor catalysts for the majority of other reactions. Most of the enzymes, for example the ones involved in GSL biosynthesis, are not perfectly specific for a single substrate and often select between several substrates that are available simultaneously [95]. Therefore, enzyme specificity can be defined as how the enzymes discriminate between substrates present in the same reaction mixture.

Met-derived glucosinolates in Arabidopsis have side-chain lengths of three to eight carbon atoms, i.e. one to six additional carbon atoms [45]. These glucosinolates are referred to as 3C to 8C glucosinolates. Additionally, the accumulation of glucosinolates is evaluated in terms of accumulation of short- and long- chain glucosinolates. For mathematical convenience, we will use the terms $3Cness$ and $7Cness$ that define the relative concentrations of the short- and long- chain glucosinolates as

$$3Cness = \frac{[3C]}{[3C] + [4C]} \qquad 7Cness = \frac{[7C]}{[7C] + [8C]}. \tag{2.19}$$

To investigate the regulation of $3Cness$ and $7Cness$, we simulated the steady-states for different values of dissociation constant ($K_{d_{i,j}}$) of the reaction, where substrate of chain-length $j = 0$ is catalysed by the enzyme $i$. The reactions catalysing substrates of chain-length $j = 0$ of different metabolic enzymes are reminded here;

| Symbol | Reaction | Enzyme |
|---|---|---|
| $R_1$ | $K_{p,0} \longrightarrow K_{p,1}$ | MAM3 |
| $R_7$ | $K_{p,0} \longrightarrow K_{p,1}$ | MAM1 |
| $R_{10}$ | $K_{c,1} \longrightarrow K_{p,1}$ | BAT5 |
| $R_{16}$ | $K_{p,0} \longrightarrow K_{c,0}$ | BAT5 |
| $R_{23}$ | $A_{c,0} \longrightarrow A_{p,0}$ | AATR |
| $R_{30}$ | $A_{p,0} \longrightarrow A_{c,0}$ | AATR |
| $R_{37}$ | $A_{c,0} \longrightarrow K_{c,0}$ | BCAT3 |
| $R_{44}$ | $K_{c,0} \longrightarrow A_{c,0}$ | BCAT3 |
| $R_{51}$ | $A_{c,0} \longrightarrow K_{c,0}$ | BCAT4 |
| $R_{58}$ | $K_{c,0} \longrightarrow A_{c,0}$ | BCAT4 |
| $R_{65}$ | $A_{c,1} \longrightarrow A_{c,1}$ | CYP79F1 |

Table 2: List of reactions catalysing substrates $S_{i,j}$ of chain-length $j = 0$ by different metabolic enzymes. While *A* and *K* denote the amino and keto acid intermediates, the subscript *c* and *p* denote the sub-cellular localisation *cytosol* and *plastid*, respectively.

The change in the steady-state 3Cness (7Cness) is measured in terms of coefficients of response $C_R$, which is calculated by

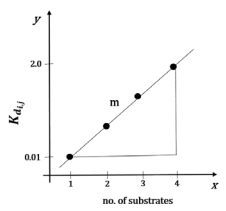
$$C_R = \frac{R' - R}{R} \tag{2.20}$$

where $R$ and $R'$ denote the 3Cness (7Cness) at $K_{d_{i,j}}$ at 1.0 and new value, respectively. Further, we assume that the $K_{d_{i,j}}(j = 0, \ldots, n)$ of the reaction with substrate of chain-length $j = n$ of an enzyme $i$ depends on the $K_{d_{i,j}}(j = 0)$ as

$$K_{d_{i,n}} = 2 - K_{d_{i,0}}, \tag{2.21}$$

where 2 (arbitrary units) is a chosen maximum of the $K_{d_{i,j}}$ values. Moreover, the $K_{d_{i,j}}$ depends on the dissociation constant of the reaction catalysing substrate with $j - 1$ as

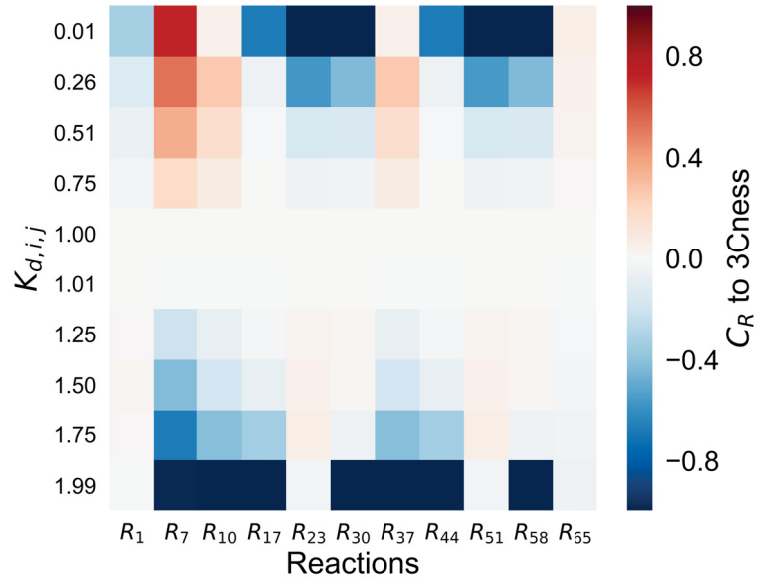
$$K_{d_{i,j}} = m \times K_{d_{i,j-1}}. \tag{2.22}$$



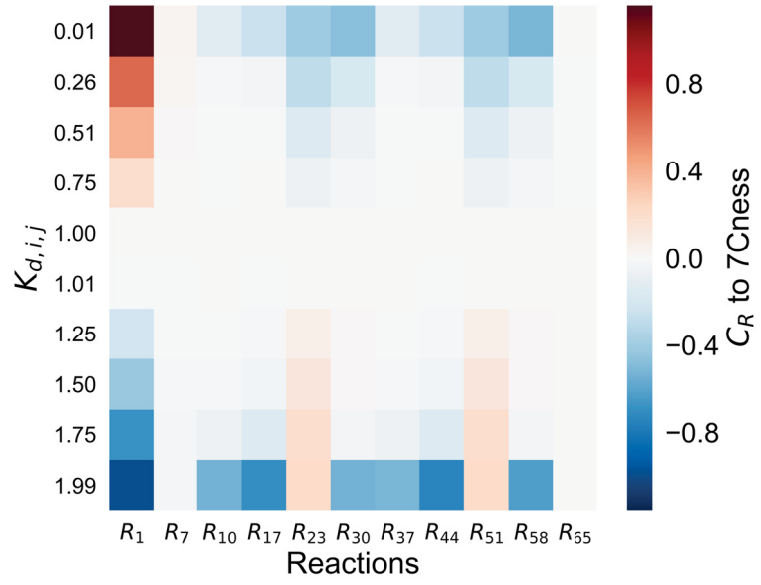

*where m is the slope caluclated as* $m = \Delta x/\Delta y$.

In Eq. 2.22, $m$ is a constant, which is calculated as a ratio between

the change in $K_{d_{i,j}}$ values to the number of substrates that can be catalysed by the enzyme $i$.




(a)




(b)

Figure 12: Heatmap showing the coefficients of response ($C_R$) to 3Cness and 7Cness at different $K_{d_{i,j}}$ of reaction with substrate of chain-length $j = 0$ catalysed by enzyme $i$. Figure (a) shows the heatmap of $C_R$ to 3Cness, whereas figure (b) presents the heatmap of $C_R$ to 7Cness. Different values of $K_{d_{i,j}}$ is presented on *y-axis*. Whereas, along the *x-axis* symbols of reactions catalysing substrate with $j = 0$ are given. Details on the reaction symbols are given in Table 2. Results are based on simulations performed using *vanilla* parameters (Table 4).

Figure. 12 shows heatmaps of the coefficients of response to (a) 3Cness and (b) 7Cness calculated for different $K_{d_{i,j}}$ of reactions catalysing substrate with chain-length $j = 0$ of enzyme $i$. Fig. 12a shows that a decrease in $K_{d_{i,0}}$ of $R_7$ catalysed by MAM1 results in high $C_R$ value. According to Eq. 2.21 and 2.9, low $K_{d_{i,0}}$ results in faster rate of reaction that catalyses $K_{p,0}$ than the reaction catalysing $K_{p,1}$. Consequently, the flux is directed towards more production of 3C GSLs, thus increasing 3Cness. However, a high $K_{d_{i,0}}$ results in slower rate of the reaction that catalyses $K_{p,0}$ than reaction catalysing $K_{p,1}$. Thus, the flux is diverted towards the production of longer chain GSLs, and thus decreasing the 3Cness. Varying the $K_{d_{i,0}}$ of reaction $R_1$, where MAM3 catalyses $K_{p,0}$, negatively regulates (negative coefficient values) the 3Cness. Further, variation in the $K_{d_{i,0}}$ of BAT5-mediated transport reactions ($R_{10}$), shows an increase in 3Cness at low $K_{d_{i,0}}$ and *vice-versa*. Whereas, the other transport reactions ($R_{17}, R_{23}, R_{30}$) show a negative regulation of 3Cness if the $K_{d_{i,0}}$ is varied. A similar behaviour is seen if the $K_{d_{i,0}}$ values of the transamination reactions ($R_{37}, R_{44}, R_{51}, R_{58}$) catalysed BCAT3 and BCAT4 are varied. Lastly, a mariginal change in the 3Cness is observed if the $K_{d_{i,0}}$ of reaction $R_{65}$ catalysed by CYP79F1 is varied. Similarly, we studied the effect of varying $K_{d_{i,0}}$ on the 7Cness of stead-state flux. Figure 12b shows a heatmap of the coefficients of response ($C_R$) to 7Cness. Decreasing the $K_{d_{i,0}}$ of the MAM3-catalysed reaction $R_1$, increases the *7Cness*. As per Eq. 2.22, the $K_{d_{i,0}}$ of reaction catalysing $K_{p,4}$ is smaller than $K_{p,5}$, which results in the faster kinetics leading to the diversion of flux towards the production of 7C glucosinolates. Thus, increasing the 7Cness. However, increasing the $K_{d_{i,0}}$ results in the diversion of flux towards the 8C glucosinolates, and thus decreasing the 7Cness. Conversely, the reactions $R_{23}$ and $R_{51}$ increase the 7Cness with increasing value of $K_{d_{i,0}}$, and *vice-versa*. Whereas, the other reactions ($R_{10}, R_{17}, R_{30}, R_{37}, R_{44}, R_{58}$) show a marginal decrease in the 7Cness if $K_{d_{i,0}}$ is varied. However, no

change in the 7C*ness* is observed if the $K_{d_{i,0}}$ of reaction $R_{65}$ catalysed by CYP79F1 is varied.

## 2.6 DISCUSSION

The structural diversity of aliphatic glucosinolates arises from the differential elongation of its side-chain. Moreover, the chain-elongation pathway is an interplay of various enzyme isoforms that are localised in different cellular compartments. The corresponding metabolic network is so complicated that mathematical modelling is required to try to understand it. Based on our knowledge of the pathway, kinetic properties of the metabolic enzymes and some simplifying assumptions, we developed a mathematical model of the chain-elongation of Met-derived glucosinolates in *Arabidopsis thaliana*.

The presented model is based on a similar idea conceived by Knoke et al., 2009 [57] concerning the chain-length distribution of Met-derived glucosinolates. We extend the model capabilities, by employing rate laws that account for the broad-range substrate specificity of metabolic enzymes. By providing the mathematical description, we illustrate how different biosynthetic rates are affected by all metabolite concentrations, a behaviour originating from broad-range substrate specificity of the metabolic enzymes.

Using a reference set of kinetic parameters, we showcase an important phenomenon of enzyme-catalysed reactions, i.e. the overloading of substrates. Model simulations showed that at high concentration of methionine influx, the chain-elongation pathway became overloaded, and high accumulation of intermediate metabolites is observed. This may result in unusual characteristics. A biochemical example of the substrate overloading has been shown in a previous study [21], which showed how fatty acid $\beta - oxidation$ becomes vulnerable to substrate overload. Such kind of studies are an ef-

fort to estimate the critical flux that can be handled by a metabolic pathway and can produce desired concentrations of biosynthetic products. Furthermore, model simulations showed that at high substrate concentration the chain-length distribution shifts toward the production of long-chain GSLs. A plausible explanation for this behaviour is that at high Met concentrations, the transamination reaction is favoured in the direction, which results in the production of keto acids. As keto acids are direct substrates of chain-elongation, high availability of substrates leads to high production of chain-elongated products. Thus, the chain-length distribution shifts toward the production of long-chain GSLs.

Different sub-cellular localisations of metabolic enzymes make the chain-elongation pathway highly compartmentalised [30, 36, 92]. Knock-out studies revealed that the activity of bile-acid transporter 5 (BAT5) as the only transporter in the biosynthesis of Met-derived glucosinolates is not true, and the presence of an unidentified yet amino acid transporter is highly speculated [30]. Model simulations showed that for very fast reversible transport reactions, the metabolite concentrations in different cellular compartments attain equilibrium. Thus, the steady-state chain-length distribution is not limited by the compartmentalisation of the glucosinolate biosynthesis. In such cases, skeleton (referred to as *generic* in this chapter) models that consist of fewer reactions can be used to reduce modelling complexities. Skeleton models have turned out to be useful in previous studies [42, 57]. Furthermore, a fine-scale investigation showed that for slow transport kinetics the steady-state biosynthetic flux is diverted towards the production of long-chain GSLs (see Fig. 10). A plausible explanation is the increase in residence time of the substrates, capable of undergoing elongation, in the scope of elongation. However, the quantitative difference (cf. Fig. 10a and 10b) in the steady-state chain-length distribution for slow kinetics of BAT5 and AATR is observed. It originates from the fact that BAT5 exerts

activity on the keto acid intermediates, whereas AATR mediates the transport of amino acid intermediates. While the chain-elongated amino acid intermediates are direct substrates that can lead to the production of GSLs, the keto acid intermediates needs to undergo a transamination reaction before it enters downstream steps of GSL biosynthesis (cf. Fig. 4). Thus, the chain-length distribution of Met-derived GSLs is also limited by the transamination reaction, which further increases the residence time of the substrates within the scope of further elongation.

Differential expression of metabolic genes could result in different concentrations of respective enzymes. Model simulations showed that the chain-length distribution is highly regulated by the concentrations of different metabolic enzymes (cf. Fig. 11). The reversal in 3C to 4C ratio seems to be associated with the activity of MAM1 enzyme (cf. Fig. 11a). Since MAM1 exerts activity on MTOB and MTOP only, high concentrations of MAM1 enzyme will increase the $V_{max}$ of the respective reactions. This results in the diversion of metabolic flux towards the production of 4C GSLs, given $V_{max}$ of all other reactions constant. A similar behaviour is exhibited at high concentration of MAM3 enzyme, which catalyses keto acids of variable chain-lengths ($K_{p,0}, \ldots, K_{p,5}$, cf. Fig. 5). Thus, at high MAM3 concentrations the chain-length distribution favours high production of 8C GSLs. Varying the concentrations of transaminases show an interesting phenomenon. While at low enzyme concentrations the chain-length distribution shifts toward longer chain GSLs, high transaminase concentrations lead to the production of short-chain GSLs. This behaviour stems from the fact that low tranaminase concentration implies low $V_{max}$ of respective transamination reactions, which in turn increases the residence time of the substrates capable of elongation within the scope of further elongation, *vice-versa*. Lastly, variation in the concentrations of CYP79 enzymes show a typical behaviour. The low concentration of CYP79F1 enzyme di-

verts the flux towards reactions catalysed by CYP79F2, which catalyses the substrates $A_{c,5}$ and $A_{c,6}$ (cf. Fig. 5). Thus, producing high concentrations of 7C and 8C glucosinolates. Whereas, the chain-length distribution is not limited by high concentrations of CYP79F1. However, an almost no effect on the chain-length distribution during the variation in CYP79F2 concentration arises from the compensatory activity of CYP79F1 enzyme. These analyses are important for finding the enzymes that are important, and should be amplified (supressed) for a desired pattern of glucosinolate accumulation.

Enzyme affinities play an important role in the steady-state concentrations of GSLs. By varying enzyme affinities toward substrates of different chain-lengths, model simulations allowed us to calculate the extent of regulation of steady-state chain-length distribution by different metabolic enzymes. The analyses indicate which parameters have to be engineered to produce a desired distribution toward short- and long-chain GSLs. The response analysis (cf. Fig. 12) only provides reliable predictions when small changes are considered. Although we considered some simplifying assumptions, model simulations provide insights and explain the systemic behaviour of the chain-elongation of Met-derived GSLs. Such analyses are important to investigate, moreover to understand, the functioning of enzymes-catalysed biosynthetic network. A practical application of such analyses is to detect which enzymes should be amplified by genetic manipulation to give the highest effect in increasing the biosynthesis of a target product.

The presented mathematical model is developed for a certain pragmatic purpose. We intend to give a detailed mathematical representation of the underlying reactions catalysed by broad-range substrate specific enzymes, which will be very important for fitting experimental data in the best possible way.

# 3

# MATHEMATICAL MODELLING OF THE CHAIN-LENGTH DISTRIBUTION OF ALIPHATIC GLUCOSINOLATES IN ARABIDOPSIS ECOTYPES

*It doesn't make a difference how beautiful the guess is. It doesn't make a difference how smart you are, who made the guess, or what his name is. If it disagrees with experiments, it's wrong.*

— Richard P. Feynman

Aliphatic glucosinolates are a major class of glucosinolates in *Arabidopsis thaliana*. These are metabolically derived from methionine (Met) [36]. Diversity of Met-derived glucosinolates depends on the variation in the chain-length and patterns of secondary oxidation and esterification [52, 99]. Glucosinolates of six variable chain-lengths, referred to as 3C to 8C glucosinolates, are found in *A. thaliana* [32]. Although, the genes involved in the biosynthesis have been identified, but much more of discovery awaits before we fully understand how certain glucosinolates with a particular pattern are produced. To address the long-standing question about the major regulators of metabolic fluxes through the biosynthetic pathway, we use our mathematical model (described in chapter 2) to reproduce patterns of glucosinolate accumulation in the leaves of *A. thaliana* Columbia (Col) ecotype. The model uses pathway architecture and experimentally measured enzyme kinetics data to evaluate the role of different metabolic enzymes/transporters in determining the diversity of Met-derived glucosinolates in *Col* leaves. However, the missing kinetic parameters are estimated using genetic algorithm (GA) [46], which is based on evolutionary concepts such as selection, reproduction, crossover and mutation. In its original form, GA

optimises the value(s) of the decision variables that would minimise a given objective function. The generic steps involved in GA are: (i) *random initialisation* (generation of random individuals to form initial population), (ii) *selection* (selection of individuals from the population based on some fitness measure), (iii) *reproduction* (generation of new individuals to maintain the population size), (iv) *crossover* (generation of new individuals by mutually exchanging the stochastic self-contained parts of the existing individuals), and (v) *mutation* (diversification of existing individuals by making a stochastic change in the population). Steps (ii-v) are repetitively performed to minimise a given objective function.

In the present chapter, we showcase the simulated results not only for the wild-type ecotype, but also for different mutant backgrounds of *Col* ecotype. Furthermore, the model is used to reproduce experimentally measured patterns of glucosinolate concentrations found in various other Arabidopsis ecotypes. Due to the gap in our knowledge of kinetic parameters in these ecotypes, we estimated the values to make a good agreement of the model simulations with the experimentally measured concentration of glucosinolates as good as possible.

## 3.1 ESTIMATION OF THE REQUIRED INFLUX OF METHIONINE

To estimate the methionine influx required to produce a typical concentration of Met-derived glucosinolates in the Columbia ecotype of *A. thaliana*, we did a back-of-the-envelope calculation shown in figure 13. We assume that the biosynthesis of Met-derived glucosinolates scales linearly with the growth rate of a plant cell. Given that the biosynthesis involves different cellular compartments, which have different volumes, we choose to discuss the glucosinolate biosythesis per unit cell volume. Later, we multiply by a typical growth rate of a plant cell to get the turnover per unit time. Our es-

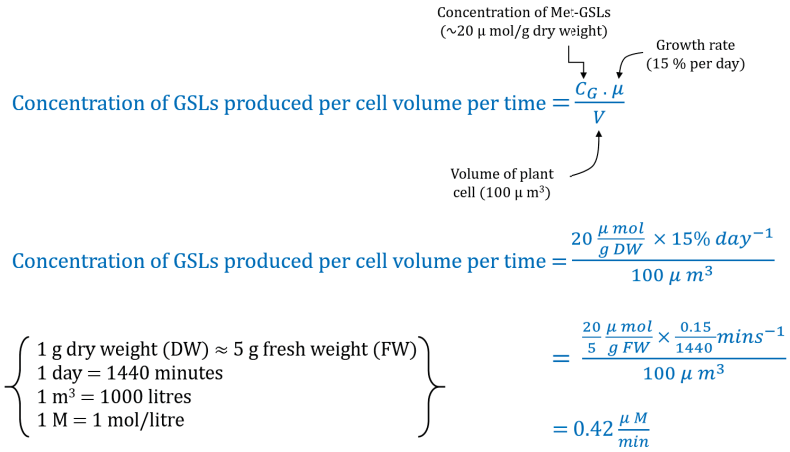



Figure 13: A back-of-the-envelope calculation of the methionine flux to produce a typical concentration of Met-derived glucosinolates in *A. thaliana* leaves. Estimation is based on generic parameter values [55, 75, 92, 97].

timate relies on *generic* values of the parameters. The typical concentration of Met-derived glucosinolates is taken from [92]. A typical growth rate and volume of plant cell are taken from the BioNumbers database (cf. [75], also see BNID: 107725, 108685). From our calculations, an influx of 0.42 $\mu M \cdot min^{-1}$ is required to produce a typical concentration (20 $\mu mol \cdot g^{-1}$ dry weight) [55, 92, 97] of Met-derived GSLs in the *Col* ecotype of *A. thaliana*.

## 3.2 KINETIC PARAMETERS

The presence of glucosinolates in the model plant, *Arabidopsis thaliana*, has boosted the glucosinolate research significantly, resulting in kinetic studies of various metabolic enzymes. Our model constitutes the activity of eight enzymes/transporters that correspond to 144 kinetic parameters, of which 25 values are known [10, 55, 58, 92, 97]. However, the measured values are based on *in vitro* assays, thus using them directly for quantitative modelling is not straightforward. $V_{max}$, for example, is proportional to the enzyme concentrations that is not known for any of the enzymes involved in chain-elongation. As a rough estimation, protein concentration scales lin-

early with cell mass and volume. Given cell volume can change several folds as a consequence of growth, it seems reasonable to discuss the $V_{max}$ values per unit volume. Thus, we rescale all experimentally measured $V_{max}$ such that the virtual concentration of each enzyme in the wild-type is 1 $mg^{-1} \cdot l^{-1}$; $V'_{max} = V_{max}/mg^{-1} \cdot l^{-1}$. A similar approach has been used in [57]. Thus, the rescaled $V_{max}$ is denoted as $V'_{max}$ in Table 4.

## 3.3 ESTIMATION OF THE MISSING KINETIC PARAMETER VALUES

Our model constitutes 144 kinetic parameters, of which 119 values are yet unknown. Thus, the missing kinetic values are estimated, so as to optimise the agreement between the model-simulated and the experimentally-measured Met-derived glucosinolate (GSL) concentrations in the leaves of *A. thaliana Col* ecotype. GA is used to estimate the missing values of kinetic parameters. By an iterative process, GA optimises the parameter values to minimise the following objective function:

$$\Delta = \| G - \hat{G} \| \tag{3.1}$$

where, $\Delta$ corresponds to the euclidean distance between the experimentally-measured GSL concentration vector G ($G = 3C, \ldots, 8C$) and model simulated steady-state flux vector $\hat{G}$ ($\hat{G} = v_{G,1}, \ldots, v_{G,6}$). Since we assume that the steady-state flux is diluted by growth to produce respective GSL concentration, the flux is compared to the steady-state concentration. For the purpose of comparison, the values are normalised to the experimentally measured total concentration [92] of Met-derived GSLs and influx $v_0$ ($v_0 = 0.42\mu M \cdot min^{-1}$) for vectors G and $\hat{G}$, respectively.

GA based parametrisation does not yield just one *best* set of kinetic parameters but rather a high number of parameter sets characterised by similar agreement between the model-predicted and the experimentally-determined GSL profile. A fiducial set of kinetic parameters that reproduces the wild-type and mutant patterns of GSL accumulation in the leaves of *A. thaliana Col* ecotype is given in Table 4.

## 3.4 RESULTS

### 3.4.1 *Reproducing the pattern of glucosinolate accumulation in the Columbia ecotype*

Met-derived GSLs of six variable chain-lengths are found in *Arabidopsis thaliana*. The wild-type Columbia (Col) ecotype of *A. thaliana* accumulates high concentrations of GSLs that are four-carbon long and low concentrations of other chain-elongated glucosinolates. To reproduce the GSL pattern of wild-type *Col* ecotype, we used the fiducial parameters (Table 4) to simulate the steady-state chain-length distribution of GSLs for influx $v_0 = 0.42 \quad \mu M \cdot min^{-1}$.
Figure. 14 shows a comparison of experimentally measured GSL concentrations to the model simulated steady-state GSL concentrations of Met-derived GSLs in the leaves of *A. thaliana* Col wildtype ecotype. While the blue bars denote the experimentally measured concentrations of the different Met-derived GSLs, orange bars denote the steady-state GSL concentrations from model simulations. For the purpose of comparison, the values are normalised to the total concentration of Met-derived GSLs. Model simulation of chain-length distribution is in good agreement with the experimentally-measured GSL profile.

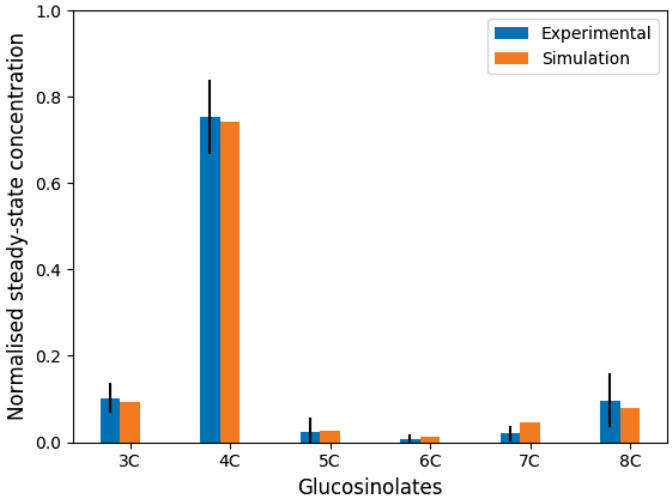



Figure 14: Comparison of the experimentally measured concentrations with the model-simulated concentrations of Met-derived glucosinolates in *A. thaliana* Col wildtype (WT) ecotype. The values are normalised to the the total concentration of Met-derived GSLs. 3C, three carbon GSLs; 4C, four carbon GSLs; 5C, five carbon GSLs; 6C, six carbon GSLs; 7C, seven carbon GSLs; 8C, eight carbon GSLs. Experimental data of GSL concentrations in the WT ecotype is taken from [92]. Simulation is performed using the parameters given in Table 4.

### 3.4.2 *Reproducing the GSL profiles of knockout mutants of Col ecotype*

From literature [95], we know that there exist different isoforms of GSL biosynthetic enzymes in *Col* ecotype. MAM synthases, for example, exist as MAM1 and MAM3 enzymes. While MAM3 accepts as substrates all six of the 2-oxo acid intermediates of chain-elongation reactions, MAM1 exerts activity only on the two shortest 2-oxo acids, MTOB and MTOP [58, 97]. Likewise, the chain elongation pathway includes several transamination reactions with Met or amino-acid derivatives and 2-oxo acids of variable side chains. In principle, BCAT4 and BCAT3, respectively, could catalyse all cytosolic and plastidic transamination reactions with differentially-elongated 2-oxo acid intermediates of the chain-elongation pathway. Lastly, there also exist two CYP79 enzymes, CYP79F1 and CYP79F2. While the former catalyses the conversion of homomethionines of

various chain-lengths to their respective aldoximes, the latter catalyses the conversion of the two longest chain homomethionines, only. To investigate the role of different enzymes (isoforms) in regulating the steady-state concentrations of GSLs, we used the kinetic parameters (Table 4) to simulate the steady-state for $v_0 = 0.42\mu M \cdot min^{-1}$ in different mutant conditions. To model the knockout, we set the maximum velocity $V_{max}$ of their catalysed reactions to zero. Moreover for the mathematical clarity, we assume that there exist no compensatory expression of other enzymes.

Figure. 15 shows a comparison of the experimentally-measured leaf GSL profile with the model-simulated profile in *Col* MAM1 and MAM3 knockout mutants. The length of the bars in Fig. 15 are

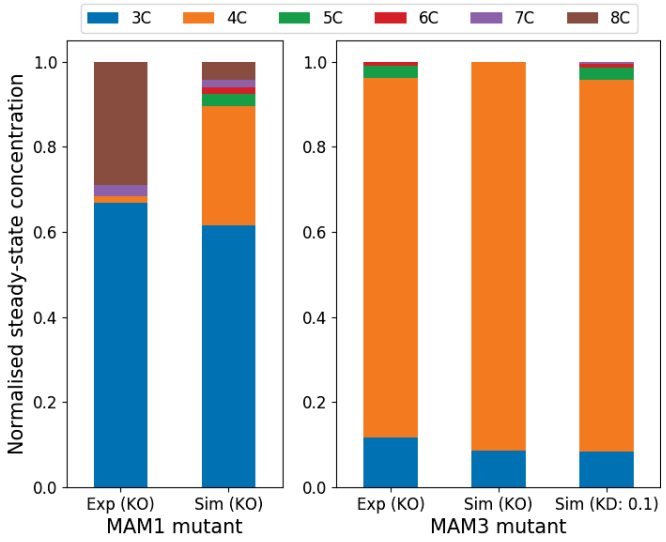



Figure 15: Comparison of the experimental (Exp) leaf GSL profile with the model simulated (Sim) GSL profile in MAM1 and MAM3 knockout (KO) and knockdown (KD) mutants of *Col* ecotype. "KD : 0.1" refers to the knockdown to 10 percent of original MAM3 concentration. The length of the bars are normalised to the total concentration of Met-derived GSLs. Results are based on simulations performed using parameters (Table 4). Experimental data on mutant profiles are taken from [58, 97].

normalised to the total concentration of Met-derived GSLs. Model simulations show a good agreement with the experimental observations. The reversal in the distribution of 3C to 4C glucosinolates attributes to the loss of MAM1 activity. Whereas, the loss of MAM3

activity leads to the diversion of flux towards the production of 3C and 4C glucosinolates only (see Fig. 15). The simulations of the MAM3 knockout mutant also showed no flux towards the production of longer-chain glucosinolates (5C . . . 8C), while low amounts of 5C and 6C glucosinolates were observed in the actual plant analyses [97]. A possible explanation of the consequence could be the residual activity of MAM3. Thus, to test our speculation we performed a knockdown to 10 percent of original MAM3 concentration. To model the knockdown of an enzyme, we set the $V_{max}$ values of their catalysed reaction to 10 percent of the original $V_{max}$ values. Figure 15 also shows a comparison of the experimentally-measured concentrations to the model simulated knockdown (KD) concentrations of different Met-derived GSLs in MAM3 mutants. Simulations show a good agreement with the experimental observations. Thus, it is plausible to assume a knockdown of MAM3 rather than a knockout.

Further, we simulated the knockout of transaminases, BCAT3 and BCAT4. Figure 16 shows a comparison of the experimentally-observed leaf GSL profiles with the simulated GSL profile of BCAT3 and BCAT4 mutants. While the simulated BCAT4 knockout mutants showed a qualitative agreement with the experimental observations, BCAT3 mutants showed a profile that cannot be explained directly. For theoretical explanation of the observed GSL profile in BCAT3 knockout mutants, we investigated different knockdown concentrations of BCAT3. Figure 16 also shows a comparison of the simulated steady-state GSL profile in the knockdown of BCAT3 enzyme. A good agreement with the experimental observations was found at knockdown to 40 percent of original BCAT3 concentration. Thus, a plausible explanation of the experimental observations is a knockdown of BCAT3 rather than a knockout.

Lastly, we simulate the steady-state GSL profiles in CYP79F1 and CYP79F2 mutants. Figure 17 shows a comparison between the

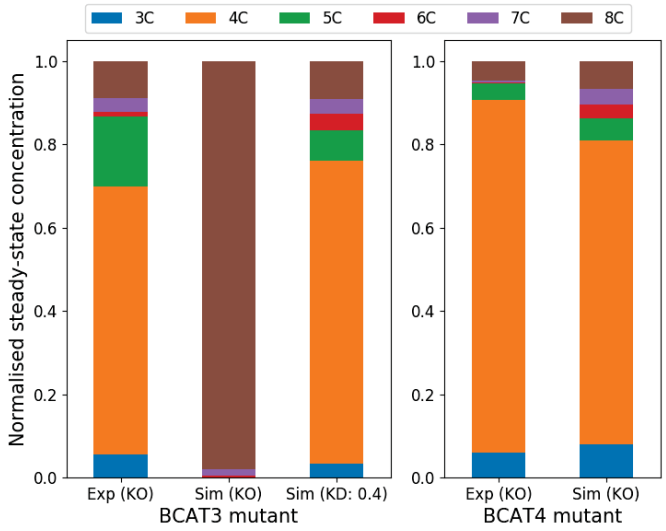



Figure 16: Comparison of the experimental (Exp) GSL profile with the model simulated (Sim) GSL profile in BCAT3 and BCAT4 knockout (KO) and knockdown (KD) mutants of *Col* ecotype. "KD : 0.4" refers to the knockdown to 40 percent of original BCAT3 concentration. Results are based on simulations performed using parameters (Table 4). Experimental data on mutant profiles are taken from [55, 92].

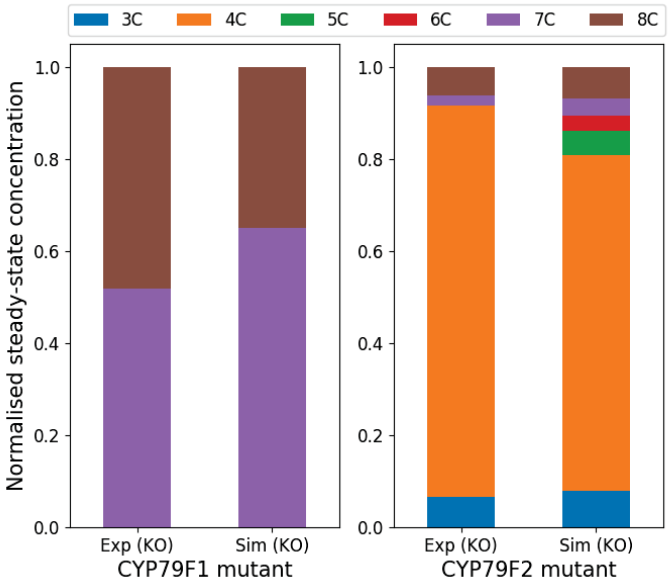



Figure 17: Comparison of the experimental (Exp) GSL profile with the model simulated (Sim) profile in CYP79F1 and CYP79F2 knockout (KO) mutants of *Col* ecotype . Results are based on simulations performed using parameters (Table 4). Experimental data on mutant profiles are taken from [10].

experimentally-measured and model simulated knockout GSL profiles in CYP79F1 and CYP79F2 mutants. Model simulations showed

a qualitatively-good agreement with the experimental observations. When we simulated the knockout of CYP79 enzymes, we were able to reproduce the qualitative distribution of the different Met-derived GSLs. However, the quantitative discrepancies may have arisen due to the compensatory expression of other metabolic enzymes.

### 3.4.3 *Metabolic control analysis of the chain-length distribution of GSLs*

From biological point of view, it is important to characterise the role of particular reactions taking place within a cell in determining the various dynamic modes of metabolism. Due to high number of variables and strong stoichiometric and regulatory relations, it seems to be impossible to gain such insights by qualitative considerations. A theoretical framework, known as *metabolic control analysis*, was developed to quantitatively explain to what extent the various reactions of a metabolic pathway determine the fluxes and concentrations [41, 50].

To investigate and quantitatively estimate the control possessed by different kinetic parameters on steady-state chain-length distribution of GSLs, we use the concepts of *control coefficients* proposed, independently, by Kacser and Burns [50], and Heinrich and Rappoport [41]. In its original form, metabolic control analysis was designed to quantify the concept of rate limitation in complex enzyme-catalysed systems. The steady-state fluxes $J_j$ in a metabolic system depend on the values of total enzyme concentrations $E_k$ [50] and other kinetic parameters [41], that affect reaction rates. Correspondingly, the *flux control coefficients* are defined as:

$$C^{J_j}_{E_k} = \left(\frac{E_k}{J_j}\frac{\Delta J_j}{\Delta E_k}\right)_{\Delta E_k \to 0} = \frac{E_k \delta J_j}{J_j \delta E_k} \tag{3.2}$$

$$C^{J_j}_{\nu_k} = \left(\frac{\nu_k}{J_j}\frac{\Delta J_j}{\Delta \nu_k}\right)_{\Delta\nu_k \to 0} = \frac{\nu_k \delta J_j}{J_j \delta \nu_k}, \tag{3.3}$$

which relates the fractional change in the steady-state fluxes to the fractional change in the total enzyme concentrations ($E_k$) and kinetic parameter ($\nu_k$) in Eq. 3.2 and 3.3, respectively. $\Delta E_k$ and $\Delta \nu_k$ denote the change in the activity of reaction *k* due to the influence of a change of the enzyme concentration and kinetic parameter, respectively. As mathematically, the fluxes $J_j$ cannot be only expressed as functions of the rates $\nu_k$, the Eq. 3.3 regards a kinetic parameter $p_k$ that affects only reaction k directly, that is,

$$\frac{\delta \nu_k}{\delta p_k} \neq 0; \qquad \frac{\delta \nu_j}{\delta p_k} = 0 \qquad \text{for any}(j \neq k). \tag{3.4}$$

Thus, Eq. 3.3 can be written as

$$C^{J_j}_{\nu_k} = \frac{\nu_k \delta J_j / \delta p_k}{J_j \delta \nu_k / \delta p_k}. \tag{3.5}$$

Thus, the coefficients defined in Eq. 3.5 can be used to interpret the extent to which reaction *k* controls the steady-state flux.
We used the aforementioned control coefficient concept to quantitatively investigate the role of different kinetic parameters in shaping the steady-state GSL profile in *Col* ecotype. Using Eq 3.5, we derive the following equation:

$$C^{R}_{p_k} = \left(\frac{p_k}{R}\frac{\Delta R}{\Delta p_k}\right)_{\Delta p_k \to 0} = \frac{p_k \delta R}{R \delta p_k}, \tag{3.6}$$

where *R* denotes the 3Cness (7Cness),

$$3Cness = \frac{[3C]}{[3C]+[4C]} \qquad 7Cness = \frac{[7C]}{[7C]+[8C]}, \tag{3.7}$$

of a steady-state GSL profile of *Col* ecotype. Thus, the Eq. 3.6 is used to investigate the control of 3Cness and 7Cness by different

kinetic parameters of enzyme-catalysed reaction of chain-elongation pathway. Figure 18 shows the coefficients of response to *3Cness* and *7Cness* of GSL profile to a finite perturbation in the kinetic parameters. Practically, we varied the model parameters ($p_k$, in Eq. 3.6) by 1 percent and calculated the change in the 3Cness and 7Cness.

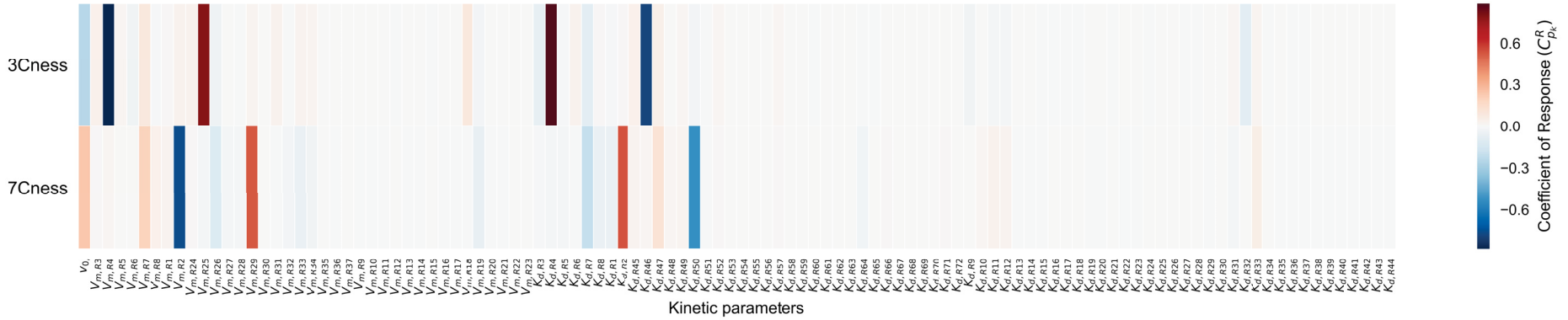



Figure 18: Heatmap of coefficients of response ($C^R_{p_k}$) to 3C*ness* and 7C*ness* of steady-state GSL profile of *Col* ecotype. Along the *x-axis* kinetic parameters p are listed, where $V_m$ and $K_d$ stand for maximum velocity and dissociation constant of enzyme-catalysed reaction k (k = 1,...,72, cf. Table 3) of the chain-elongation pathway. $v_0$ denotes the methionine influx. Results are based on simulations performed using *Col* kinetic parameters given in Table 4. $3Cness = [3C]/([3C] + [4C])$; $7Cness = [7C]/([7C] + [8C])$.

The analysis showed that the methionine influx $v_0$ plays an important role in controlling the chain-length distribution. While it negatively regulates (negative coefficients) the steady-state 3C*ness*, it positively regulates (positive coefficients) the 7C*ness*. It is understandable in a way that an increase in the substrate concentration will accelerate all the reactions producing substrates for further chain-elongation. Activity of MAM synthases, *MAM1* and *MAM3*, have a higher control (coefficient) on the steady-state 3C*ness* and 7C*ness* of the GSL profile. While the $V_{max}$ of *MAM1* positively regulates the 3C*ness*, the $K_d$ of *MAM1* negatively regulates the 7C*ness*. Whereas, activity of *MAM3* regulates both 3C*ness* and 7C*ness* of the GSL profile.

### 3.4.4 *Simulating the GSL profiles from different Arabidopsis ecotypes*

Depending upon the herbivores, several glucosinolates can act as feeding deterrents or stimulants [36]. One possible effect of this heterogeneous natural selection on glucosinolates is the rapid evolution of new GSLs or new patterns GSL accumulation. If we assume that the glucosinolate biosynthesis pathway is conserved across different Arabidopis ecotypes, we can use our model to simulate the steady-state GSL concentrations in different ecotypes.

In this section, we present the results of model simulations to reproduce different patterns of glucosinolate accumulation that are reported in a study by Kliebenstein et al., 2001 [52]. For the purpose of this study, we have selected a few ecotypes, namely *Pi*-0, *Cvi*, *Aa*-0, and *Mt*-0, that show diverse patterns of GSL concentrations. While *Pi-o* ecotype exhibits high accumulation of 3C glucosinolates, *Cvi* ecotype exhibits an intermediate distribution of 3C and 4C glucosinolates, and *Aa*—0 and *Mt*—0 ecotypes exhibit high accumulation of 4C glucosinolates. The unavailability of kinetic studies in these ecotypes leaves a gap in our knowledge of the kinetic param-

eters in these ecotypes. Nevertheless, we use the genetic algorithm to estimate kinetic parameter values to individually find a good agreement between the model simulated and the target glucosinolate profiles.

Figure 19 shows a comparison between the experimentally-measured and model simulated steady-state concentrations of Met-derived GSLs from the leaves of *Pi*-0, *Cvi*, *Aa*-0, and *Mt*-0 wild-type ecotypes. The length of the bars in Fig. 19 are normalised to the total

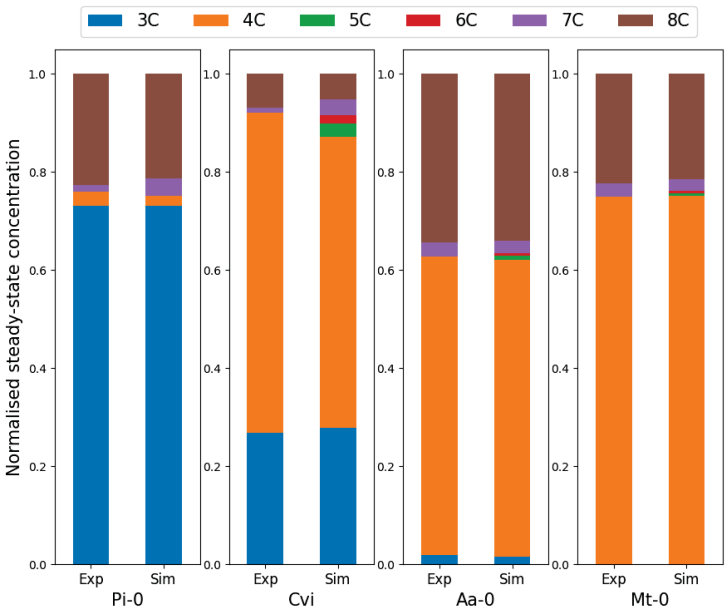



Figure 19: Comparison of the experimentally-measured (Exp) and model simulated (Sim) steady-state concentrations of Met-derived GSLs from *Pi*—0, *Cvi*, *Aa*—0, and *Mt*—0 wild-type Arabidopsis ecotypes. The experimental data of GSL concentrations in the wild-type ecotypes are taken from [52]. Results are based on simulations performed using the kinetic parameters given in Table 5

concentrations of Met-derived GSLs in the ecotypes. Model simulations show a good agreement between between the experimentally-measured and model simulated GSL concentrations in the ecotypes. Simulations are performed using the kinetic parameters given in Table 5.

The chain-length distribution is highly variable in *Arabidopsis thaliana* species. Most ecotypes accumulate glucosinolates with either three carbon (3C) or four carbon (4C) molecules. In *A. thaliana*, the production of 3C and 4C glucosinolates is largely controlled by the MAM synthases, MAM1, MAM2 and MAM3 [58, 59, 97]. Interestingly enough, majority of conclusions on the accumulation 3C versus 4C glucosinolates are based on the guilt-by-association of genes. However, the mechanistic understanding of the biosynthesis as an interplay of different metabolic enzymes is poorly understood. In this chapter, we used our mathematical model (cf. Chapter 2) to investigate the regulators of chain-elongation of Met-derived glucosinolates in *A. thaliana* ecotypes.

Influx of metabolites plays an important role in governing the steady-state concentrations in an enzymatic system [41]. Based on generic values of parameters [75] in this study, we estimated the required influx ($v_0 = 0.42\mu M \cdot min^{-1}$) to produce a typical steady-state concentration ($20\mu mol \cdot g^{-1}DW$) [55, 92, 97] of Met-derived glucosinolates in the leaves of *A. thaliana* Columbia (Col) wild-type ecotype.

Our model constitutes 144 kinetic parameter values, of which 119 are not known yet. To bridge the gap in our knowledge of kinetic parameters, we used the genetic algorithm (GA) to estimate the values of missing kinetic parameters to find a good agreement between the experimental observations and model simulations. Since GA performs a stochastic optimisation of parameter value(s), it yielded not just one best set of parameter values but a number of parameter sets that can be simulated to reproduce equally-good agreement with the target GSL profile. However, we selected the parameter values (Table 4) that showed a good agreement between the experimental observation and model simulation in *Col* wild-type and mutants.

With the adjusted set of kinetic parameters, we were able to reproduce the steady-state GSL concentrations in the leaves of *Col* wild-type ecotype (see Fig. 14). The good agreement of the model simulations with experimental observation suggests that the parameter values are a good estimation of the *in-vivo* kinetic properties of the metabolic enzymes, given the pathway structure and model assumptions.

We were also able to simulate different knockout mutants of *Col* ecotype. Model simulations showed how the steady-state GSL concentrations are changed as a consequence of enzyme knockouts, given the expression of other metabolic enzymes are constant in mutants. The reversal in the 3C/4C distribution correlated with the loss of MAM1 activity. However, the elevated levels of 8C GSLs, as reported in [58], cannot be explained directly. Whereas, the shift towards only production of 3C and 4C GSLs is attributed to the activity of MAM3. However, a small production of long-chain GSLs, as reported in [97], is associated to the residual activity of MAM3. The aforementioned speculation is tested in our model, and the result is shown in Fig. 15, where we also showcase the steady-state GSL concentrations for a knockdown of MAM3 to 10 percent of the original MAM3 concentration. The analysis allowed us to dissect the behaviour originating from knockdown of MAM3 rather than a knockout. When we simulated the knockouts of the cytosolic BCAT4 and the plastidic BCAT3 enzymes, a qualitative agreement and difference were observed during BCAT4 and BCAT3 knockouts, respectively. Model simulations showed that in BCAT3 knockout mutants, the chain-length distribution of GSL concentrations shifts towards the longer chain GSLs (cf. Fig. 16). It originates from the fact that a loss of the only plastidic transaminase, BCAT3, will increase the residence time of the substrates capable of undergoing chain elongation within the scope of further elongation. Thus, enabling the production of GSLs with longer chains. However, a

plausible explanation of the experimental observations may be the residual activity of BCAT3 enzyme. We tested the speculation using our model. Figure 16 also shows a simulated GSL profile for a knockdown of BCAT3 to 40 percent of the original BCAT3 concentration. The good agreement between the experimental observations and model simulations allows us to propose, in quantitative terms, that the experimental GSL profile in BCAT3 knockout mutants may arise from the knockdown of BCAT3 rather than a knockout. However, the presence of another plastidic aminotransferase, also stated in [55], which can partially maintain the biosynthetic flux can also be speculated. Lastly, we were able to simulate the GSL profiles of CYP79F1 and CYP79F2 mutants. Model simulations showed a good agreement with the experimental observations. The principal focus of the analyses was to investigate the role of enzyme knockouts in carving the chain-length distribution of Met-derived glucosinolates in *Col* ecotypes. Moreover, we dissect the behaviour originating from a knockdown rather than a knockout of metabolic enzymes.

*Metabolic control analysis* provides a theoretical framework that shows how understanding the role of different enzymes in an enzyme-catalysed reactions network is mainly achieved by measuring the changes around the *in-vivo* state after perturbations, rather than by only determining the state itself. The analysis indicates which quantities have to be measured to determine the response of a metabolic system. In this chapter, we have used the concepts from the metabolic control analysis to characterise the regulation of the chain-length distribution of Met-derived GSLs, in terms of 3C*ness* and 7C*ness*. We were able to show, in quantitative terms, that the influx $v_0$, activity of MAM1, MAM3, BCAT3 enzymes and BAT5 transporter have strong control (high coefficient values) on the 3C*ness* and 7C*ness* of GSL profile (cf. Fig. 18) of wild-type *Col* ecotype. Although control analysis only provides reliable predictions when small changes are considered, this may not be suf-

ficient for estimating a global effect by which various reactions of chain-elongation pathway could regulate the fluxes and concentrations. Nevertheless, an important practical application of metabolic control analysis is reminded here. Often, in glucosinolate research, one is interested in suppressing the metabolic activity in different biotic (abiotic) stress environments. In such cases, it is important to know the enzymes with highest control coefficients. Consequently, one can derive estimates for the gain in the desired output when the enzymes are altered in concentration or kinetic parameters. In this way, metabolic control analysis may provide tools for pathway engineering.

In the pre-final section of this chapter, we present how the model parameters are fitted to individually reproduce different patterns of glucosinolate accumulation observed in *A. thaliana* ecotypes. For the purpose, we selected four *A. thaliana* ecotypes, namely *Pi*−0, *Cvi*, *Aa*−0, and *Mt*−0, which exhibit different patterns of Met-derived GSL concentrations. With the adjusted set of model parameters, we were able to reproduce the steady-state concentrations of different Met-derived GSLs in these ecotypes. The good agreement between the experimental observations and model simulations shows result which is two folds. First, the model parameter values are a good estimate of *in vivo* kinetics of metabolic enzymes. Second, the model parameters can be easily fitted to reproduce any desired pattern of GSL concentrations. Practically, optimising a parameter space that constitutes 144 values to reproduce a glucosinolate profile that composes six concentration values is not challenging, given no constraints on parametrisation.

Natural phenotypic variation within or between species is regulated by complex networks of genes and associated polymorphisms [26, 36, 52, 66]. However, less is known about how the genetic differences shape the metabolic diversity. Glucosinolate metabolism is an excellent example that exhibits within and between species natural

variation in the accumulation of glocosinolates [36, 52]. Although much has been learned in recent years, still there exists a gap in our knowledge of understanding why and how genetic differences lead to synthesising particular glucosinolates with a fixed frequency.

Part III

# INVESTIGATING THE GENOTYPIC-PHENOTYPIC LINK

# 4

# INVESTIGATING THE NATURAL VARIATION IN THE ACCUMULATION OF GLUCOSINOLATES ACROSS *ARABIDOPSIS THALIANA* ECOTYPES

Depending upon the herbivore, specific glucosinolates (GSLs) can act as feeding deterrents or stimulants [8, 28, 62, 70, 74, 79]. A possible outcome of this heterogeneous natural selection on GSLs is the quick evolution of new compounds or new patterns of compound accumulation [15, 85]. New GSLs may increase resistance to herbivores that have become adapted to existing defences, whereas new patterns of GSL accumulation may provide a unique complement of defences by slowing down the counter-adaptation of herbivores. The glucosinolate defence system is one of the few systems wherein between and within species variation is being assessed at both phenotypic and causal genetic level [36, 95]. Almost all of the genes involved in the glucosinolate biosynthesis have been completely cloned, enabling us a deeper understanding of the roles of genome and gene duplication in the evolution of various glucosinolate pathways [18, 53]. Natural variation within or between species is regulated by a complex network of genes and associated polymorphisms [26, 52, 66]. These variations, however, complicate our understanding of how certain genes behave in context of a species as we often study a single genotype. Thus, understanding a gene, moreover a pathway, requires studies involving more than one ecotype.

The structural diversity of Met-derived GSLs can be explained by polymorphism at five genetic loci [52]. The locus that is responsible for determining the chain-elongation of Met-derived glucosinolates is GS-ELONG [67]. However, the constitution of GS-ELONG

is highly variable across different Arabidopsis ecotypes, moreover indel polymorphism of large regions are common [59]. In addition to MAM3 gene, present in all ecotypes, GS-ELONG harbours two more genes, namely MAM1 and MAM2. Some ecotypes constitute both, whereas others possess either of MAM1 and MAM2, or, as Ler-0 ecotype, a truncated (non-functional) MAM1 can be present in addition to MAM2 [59].

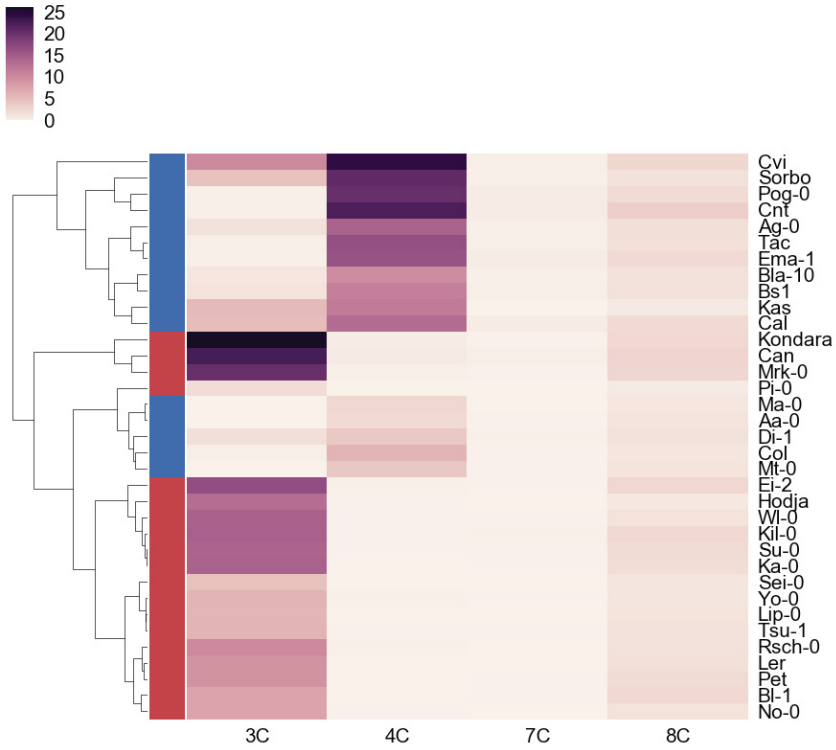



Figure 20: Clustered heatmap of the absolute concentrations of Met-derived glucosinolates (3C, . . . , 8C) in the leaves of different Arabidopsis ecotypes. The concentrations are given in $\mu mol \cdot (g \quad dryweight)^{-1}$. The red label corresponds to the ecotypes having high accumulation of 3C GSLs, whereas the blue label corresponds to the ecotypes accumulating high concentrations of 4C GSLs. See Table 10 for absolute values. 3C, three carbon GSLs; 4C, four carbon GSLs; 7C, seven carbon GSLs; 8C, eight carbon GSLs.

In a noteworthy study, Kliebenstein et. al have reported Met-derived GSL profiles from leaves of 35 different Arabidopsis ecotypes rep-

resenting a diverse sample of the geographical and environmental range of this species [52]. Figure 20 shows the concentrations of four different Met-derived GSLs, namely 3C, 4C, 7C and 8C, that are reported in [52]. It shows that the Arabidopsis ecotypes exhibit high level of diversity in both total accumulation and the chain-length distribution of Met-derived glucosinolates.

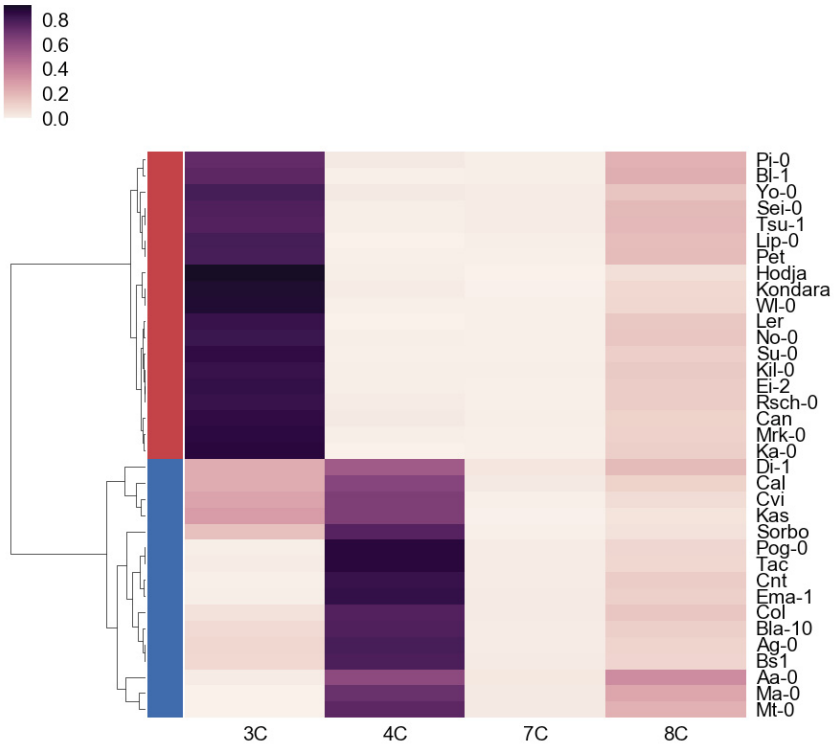



Figure 21: Clustered heatmap of the relative concentrations of Met-derived GSLs in the leaves of different Arabidopsis ecotypes. The values are normalised to the total concentration of 3C, 4C, 7C and 8C GSLs. The red coloured label corresponds to the ecotypes having high accumulation of 3C GSLs, whereas the blue label corresponds to the ecotypes accumulating high concentrations of 4C GSLs.

Furthermore in Fig. 21, we show a relative distribution of the four Met-derived GSLs, namely 3C, 4C, 7C and 8C, that are reported in [52]. It clearly shows that the Arabidopsis ecotypes either accumulate high concentrations of 3C or 4C GSLs. While the red label

corresponds to the ecotypes having high concentration of 3C GSLs, the blue label corresponds to the ecotypes accumulating high concentrations of 4C GSLs. It must be noted that no concentrations of 5C and 6C GSLs were found in these ecotypes, as reported in [52]. Thus, the normalisation will not be affected by the concentrations of 5C and 6C GSLs.

In this chapter, we investigate the link between the polymorphisms in GSL biosynthesis enzymes and their associated GSL profiles across different *A. thaliana* ecotypes. The metabolic properties of an enzyme depend on the specific order of amino acids encoded by the DNA sequence of the corresponding gene. All the enzymes have an active site, where the reactions are catalysed. Thus, the active sites compose a small number of amino acids that are essential for catalysing the reaction. Polymorphisms in the active sites of an enzyme, in principle, can change the dissociation constant ($K_d$) of the enzyme-substrate complex in an enzyme catalysed reaction. However, the gene regulatory networks may influence the maximum velocity ($V_{max}$) of respective enzyme-catalysed reaction. Thus, by relating the allelic differences to enzymatic properties, we illustrate how exclusive patterns of GSL concentrations can be produced due to the genetic variations.

## 4.1 RESULTS

### 4.1.1 *Level of polymorphism in GSL biosynthesis genes*

Natural phenotypic variation within and between species is controlled by gene regulatory networks and associated polymorphisms in metabolic genes [26, 36, 52, 66]. However, there is less clarity about the effect of polymorphisms on enzymatic properties of metabolic genes. To get an overview of the genetic diversity of GSL genes across different Arabisopsis ecotypes, we analysed the

amino acid sequences of the GSL biosynthesis genes from an example dataset constituting 343 different *A. thaliana* ecotypes taken from 1001 genome project database [1].

Figure 22 shows the diversity of GSL biosynthesis genes from 343 Arabidopsis ecotypes. The level of polymorphism of different GSL

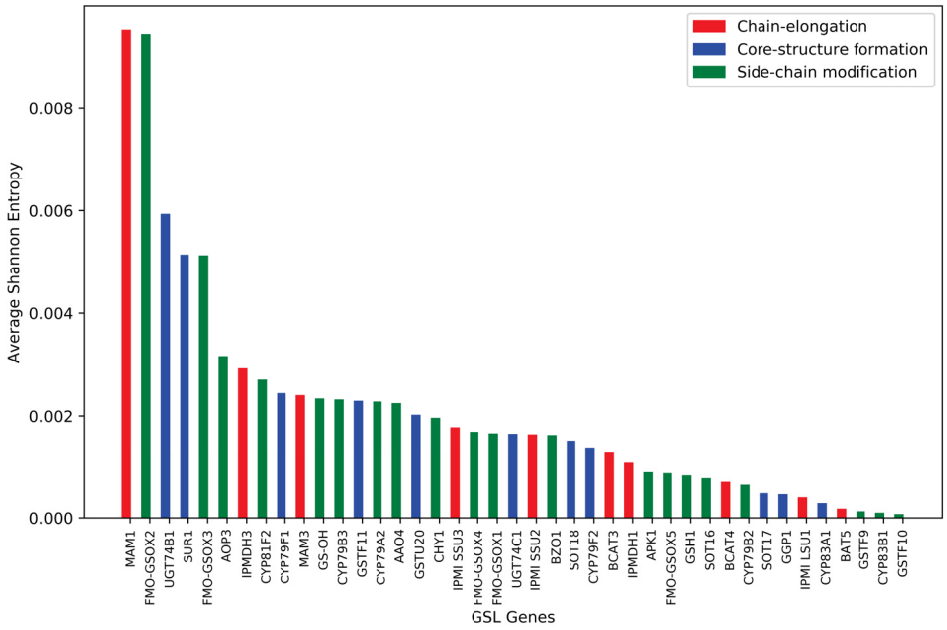



Figure 22: The level of polymorphism in the amino acid sequences of GSL genes from 343 ecotypes [1]. The bars quantify the average shannon entropy on *y-axis* across the amino acid sequence length of GSL genes plotted on x-axis. The red, blue and green bars denote entropy of genes active in the chain-elongation, core-GSL structure formation and side-chain modification of Met-derived GSLs, respectively.

genes at amino acid level is represented on *y-axis* in terms of average Shannon entropy across the gene length. Shannon entropy, $H$, associated with each data value at a position is negative logarithm of the probability of occurrence for that value [93]. It is calculated using the relation:

$$H = -\sum_{i} p_i \log p_i, \tag{4.1}$$

where $p_i$ is the probability of occurrence of a particular amino acid $i$ at a given position. For the purpose of comparison of entropy across different genes, $H$ is averaged across the gene length. Thus, the bars

in Fig. 22, represent the diversity of respective GSL gene. For visual convenience, the bars are colour coded to denote genes active in the chain-elongation process, core-structure formation and secondary chain modifications by red, blue and green colours, respectively. While we see high diversity (high entropy) exhibited by the MAM1 gene, we also see genes like BAT5 that is very conserved (low entropy) across 343 ecotypes. Moreover when the diversity of coding region was analysed, we see a different level of polymorphisms in GSL genes. Figure 23 shows the diversity in the coding regions of different GSL genes. The results cannot be explained directly but leaves a scope of further investigations of how or why variations in the coding regions are not reflected at amino acid level.

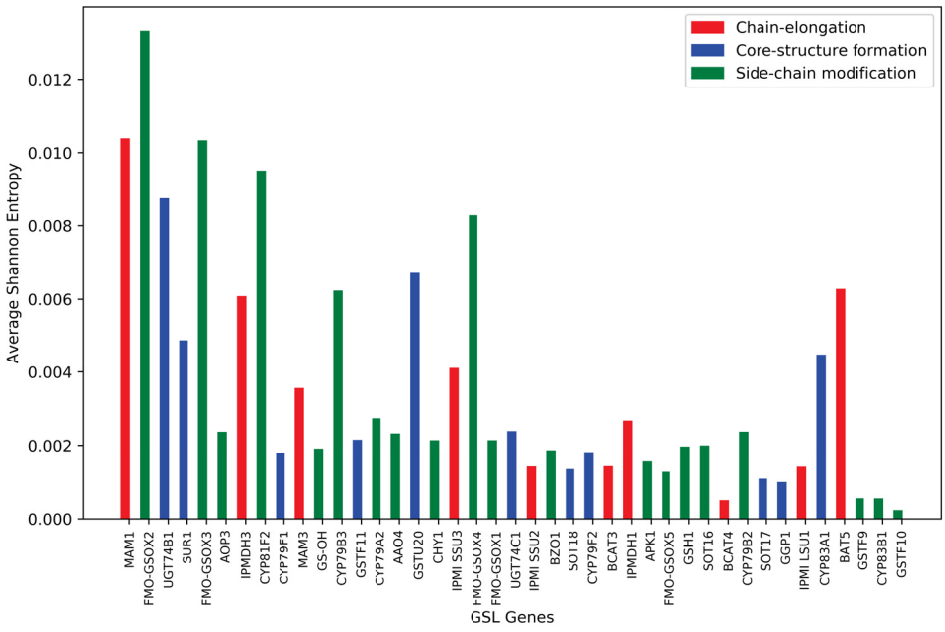



Figure 23: The level of polymorphism in the coding region GSL genes from 343 ecotypes [1]. The bars quantify the average shannon entropy on *y-axis* across the coding region length of GSL genes plotted on x-axis.

The prime focus of this study is to investigate the level of polymorphism in GSL genes at the amino acid level. Polymorphisms in the amino acid sequences can change the enzymatic properties of respective genes, given they are localised in the active sites of

the gene. However, we still have gaps in our knowledge about the presence (absence) of different GSL genes across these 343 ecotypes, before we draw any further conclusion.

### 4.1.2 *Insights into the MAM gene conundrum*

The genetic basis of chain-length distribution of Met-derived glucosinolates became evident with the identification of a locus in Arabidopsis and *Brassica napus* [67]. The locus was mapped in Arabidopsis by using a cross between two ecotypes, *Columbia (Col)* and *Landsberg erecta (Ler)*, where the major glucosinolates are 4-carbon and 3-carbon long, respectively [58]. The candidate, MAM (Methylthioalkylmalate synthase), genes are two adjacent sequences with high similarity to genes encoding isopropylmalate synthase that catalyses the condensation of chain elongation in leucine biosynthesis. Later, a third MAM-like gene, MAM2, was identified at the same locus as MAM1 [59]. The majority of Arabidopsis ecotypes examined possessed functional copies of either MAM1 or MAM2 genes. A functional MAM1 sequence has been correlated with the accumulation of 4C GSLs, whereas a functional MAM2 sequence has been correlated with the accumulation of 3C GSLs.

To gain more insights on the MAM synthases, we analysed the similarity of the annotated MAM1 gene across an example set of 343 Arabidopsis ecotypes taken from 1001 genome project database [1]. Next, we extracted the published MAM1 and MAM2 amino acid sequences from UniProt database [14]. The annotated MAM1 sequences from 343 ecotypes were compared against the extracted MAM1 and MAM2 sequences.

Figure 24 shows a mid-point rooted phylogenetic tree showing the evolutionary relationship between 343 ecotypes based on the similarities and differences to the coding region of the published MAM1 and MAM2 sequences. Based on maximum likelihood estimation

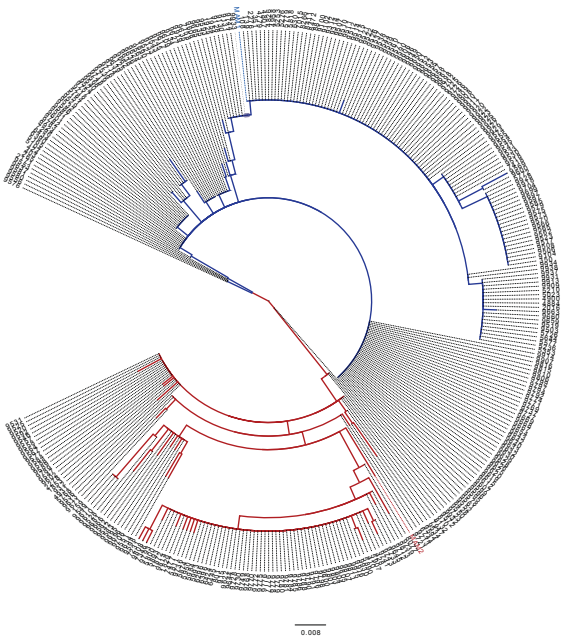



Figure 24: Mid-point rooted phylogentic tree showing the evolutionary relationship among 343 *Arabidopsis thaliana* ecotypes based similarities and differences in their annotated MAM1 amino acid sequences. While the blue branch represents ecotypes showing high similarity to the published MAM1 sequence [103], the red branch represents ecotypes showing high similarity to the published MAM2 sequence [102]. The scale bar is substitutions per position.

[34], the tree shows two main branches, clustering different ecotypes. While 226 out of 343 ecotypes, composed in the blue branch, possess high similarity to the coding region of published MAM1 gene, 117 out of 343 ecotypes in the red branch possess high similarity to the coding region of MAM2 gene. For details see Table 12. Thus, we assume that the ecotypes composed in blue and red branches possess MAM1 and MAM2 genes, respectively.

Now, we investigated the level of polymorphism in the GSL genes across ecotypes of blue and red branches by using the approach of calculating the Shannon entropy as discussed in Section 4.1.1. Figure 25 and 26 show the diversity of GSL genes across the ecotypes which are classified to possess MAM1 and MAM2 gene, respectively.

A key difference, exhibited in Fig. 25, from analysing 343 ecotypes together is the two-folds reduction in the entropy of MAM1 gene.

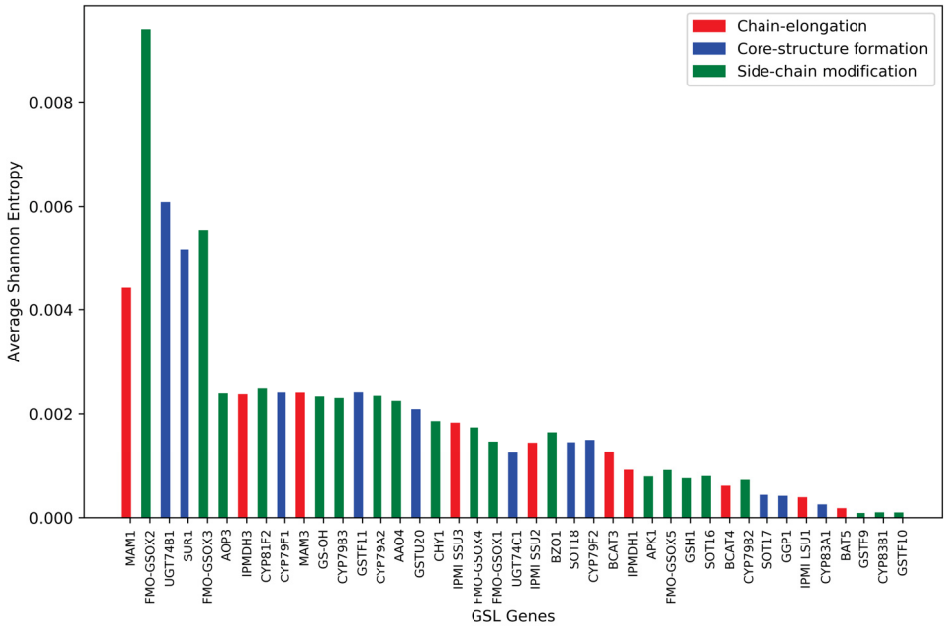



Figure 25: Diversity of GSL genes across MAM1 ecotypes from blue branch of the phylogenetic tree in Fig. 24. These ecotypes share high similarity with the amino acid sequence of published MAM1 gene [103].

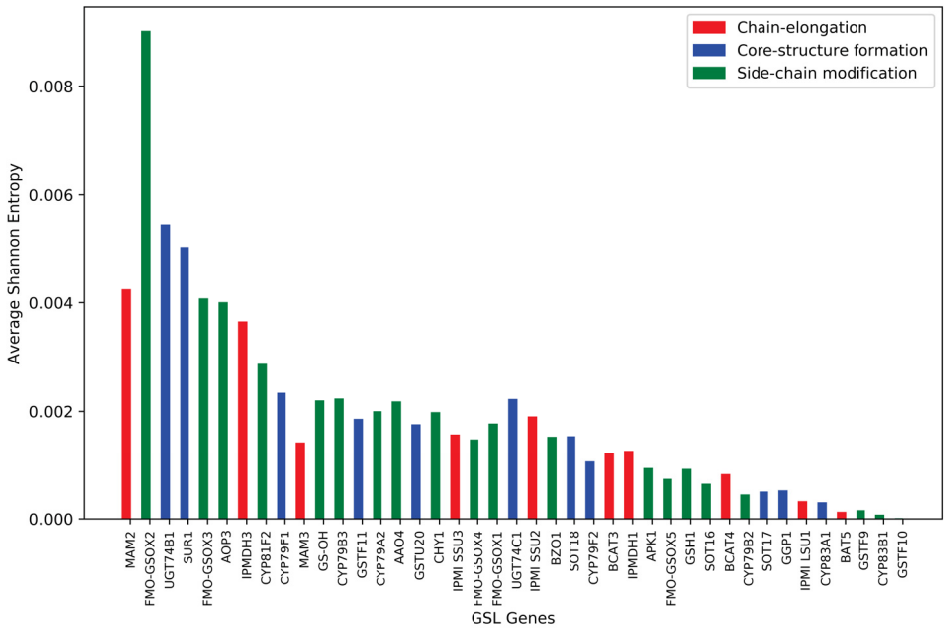



Figure 26: Diversity of GSL genes across MAM2 ecotypes from red branch of the phylogenetic tree in Fig. 24. These ecotypes share high similarity with the amino acid sequence of published MAM2 gene [102].

However, a *more-or-less* similar pattern of polymorphism as before is reflected by the other genes possesed by the MAM1 ecotypes. A similar feature, that is two folds reduction in the diversity of anno-

tated MAM1 gene is also seen in Figure 26. Since, the annotated MAM1 gene in the ecotypes of red branch shows high sequence similarity to the published MAM2 sequence, it is plausible to assume the presence of MAM2 gene rather than MAM1. The above analyses provide a general insight into the diversity of GSL genes. These analyses form the basis of dissecting the behaviour originating from presence or absence of MAM1 and MAM2 genes in the ecotypes under study.

### 4.1.3 *Analysing the metabolic genotypes and the associated metabolic phenotypes*

For practical interpretation of the genetic diversity and relating it to enzymatic properties, we studied the polymorphisms in GSL biosynthesis genes from 22 Arabidopsis ecotypes, where information about both the amino acid sequencess and the associated GSL profiles are known. These ecotypes are marked with $*$ in Table 10.
Using the approach of calculating shannon entropy, we calculated the genetic diversity of GSL genes across the 22 ecotypes. Figure 27 shows the diversity of GSL genes from 22 ecotypes. Similar to Fig. 22, the annotated MAM1 genes of the considered 22 ecotypes show high diversity. Similar to previous analyses, the published MAM1 and MAM2 sequences were used to dissect ecotypes possessing MAM1 or MAM2 genes. From our analyses, 17 out of 22 ecotypes possess MAM1 gene. Whereas five ecotypes possess MAM2 gene. See Table 10 for details.
To investigate the genetic differences across different metabolic genotypes, we calculated the genotypic distance between the metabolic genotypes of 22 ecotypes. In our study, the metabolic genotype is composed of the amino acid sequences of metabolic genes, namely *MAM1*, *MAM3*, *BCAT3*, *BCAT4*, and *BAT5*. Since, the chain-elongation pathway is highly influenced by the *CYP79*

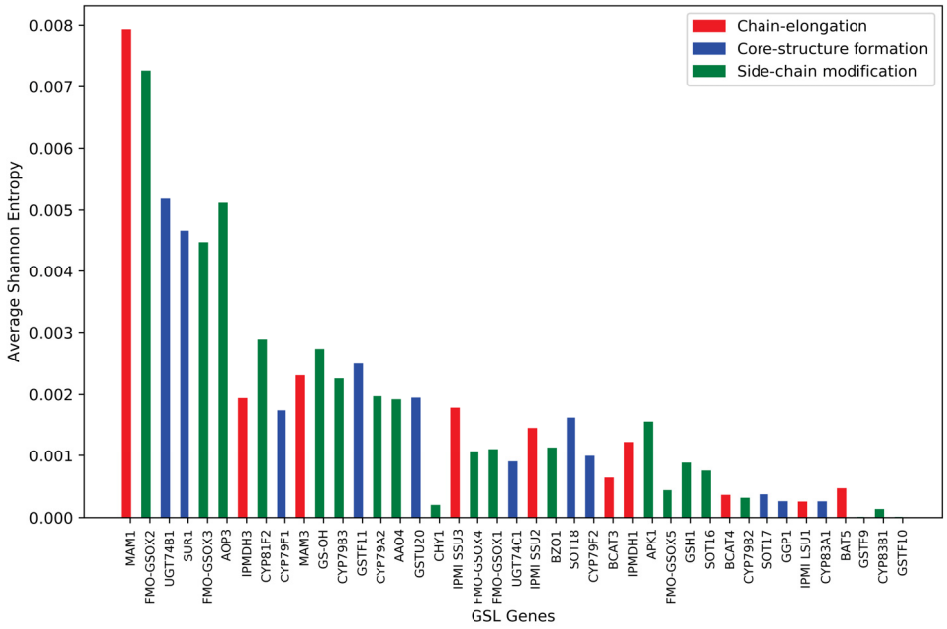



Figure 27: Diversity of GSL genes from 22 ecotypes, where information on both amino acid sequences and associated metabolic phenotypes is available [52]. The bars quantify the average shannon entropy on *y-axis* across the amino acid sequence length of GSL genes plotted on x-axis. The red, green and blue bars denote entropy of genes active in chain-elongation, core-GSL structure formation and side-chain modification, respectively.

genes [10], we included the respective amino acid sequences in the metabolic genotype. To calculate the genotypic distance we used a commonly-used measure called Hamming distance [37]. From information theory, the Hamming distance between two strings (in this case is the amino acid sequences) of equal length is the number of positions at which the corresponding characters are different. Thus, the hamming distance between metabolic genotypes $M_{G,i}(i = 1 \ldots 22)$ is calculated, such that every metabolic genotype $M_{G,i}$ gets a distance from all other genotypes. In such cases, it is useful to store the distances in a matrix. For the purpose of visualising the level of similarity of genotypes, we use multi-dimensional scaling (MDS), which is a set of ordination techniques used to visualise the information contained in a distance matrix [60]. Figure 28 shows the distances between 22 metabolic genotypes using an MDS plot. The MDS scaling of the distance matrix shows three groups

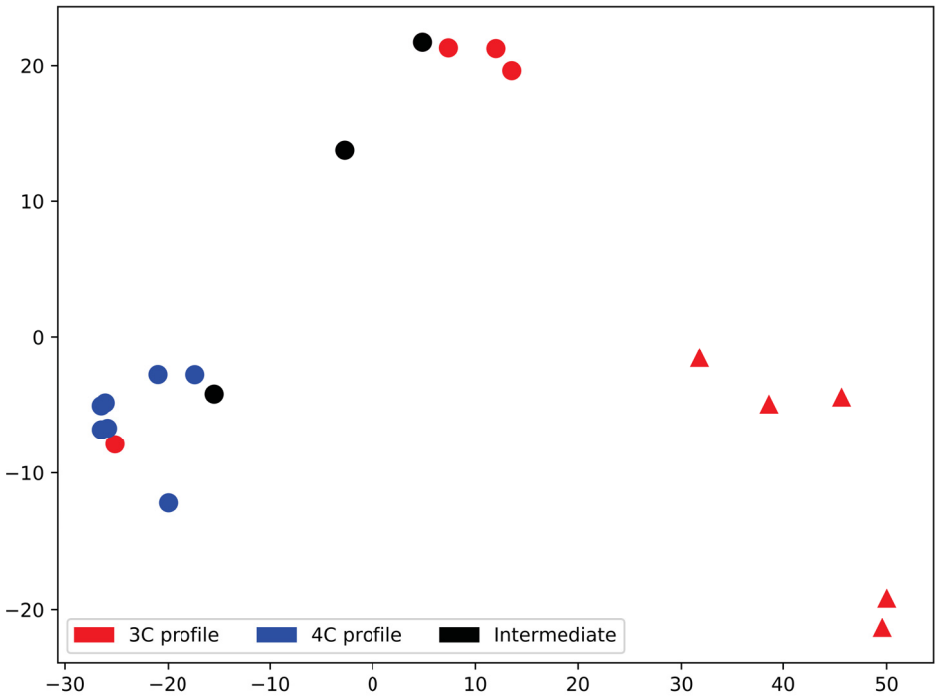



Figure 28: Multi dimensional scaling of the genotypic distance between different metabolic genotypes $M_{G,i}$. The genotypic distance is calculated in terms of hamming distance. The MAM1 ecotypes are denoted with circles, whereas the MAM2 ecotypes are represented by triangles. The ecotypes are colour coded as red and blue to represent high concentration of 3C and 4C GSLs, respectively. Whereas, the ecotypes that show an intermediate concentrations of 3C and 4C GSLs are denoted by black coloured marker.

of ecotypes. The ecotypes those show high similarity to the coding region of MAM1 are denoted by circles, whereas the ecotypes with high similarity to MAM2 are denoted by triangles in Fig. 28. The ecotypes are further classified as 3Cprofile, 4Cprofile and *intermediate*. While 3Cprofile and 4Cprofile refer to the high concentration of 3C and 4C GSLs, respectively, an *intermediate* profile refers to the intermediate concentrations of 3C and 4C GSLs. The ecotypes are colour coded as red, blue and black to denote 3Cprofile, 4Cprofile and *intermediate* profiles, respectively. A key feature seen in Fig. 28 is a clear separation of the MAM1 and MAM2 ecotypes. Also, we see one red and black circles within the group of blue circles. The results cannot be explained directly, however, investigating these

ecotypes may provide a further basis of approximating a 3C versus 4C GSL profile.

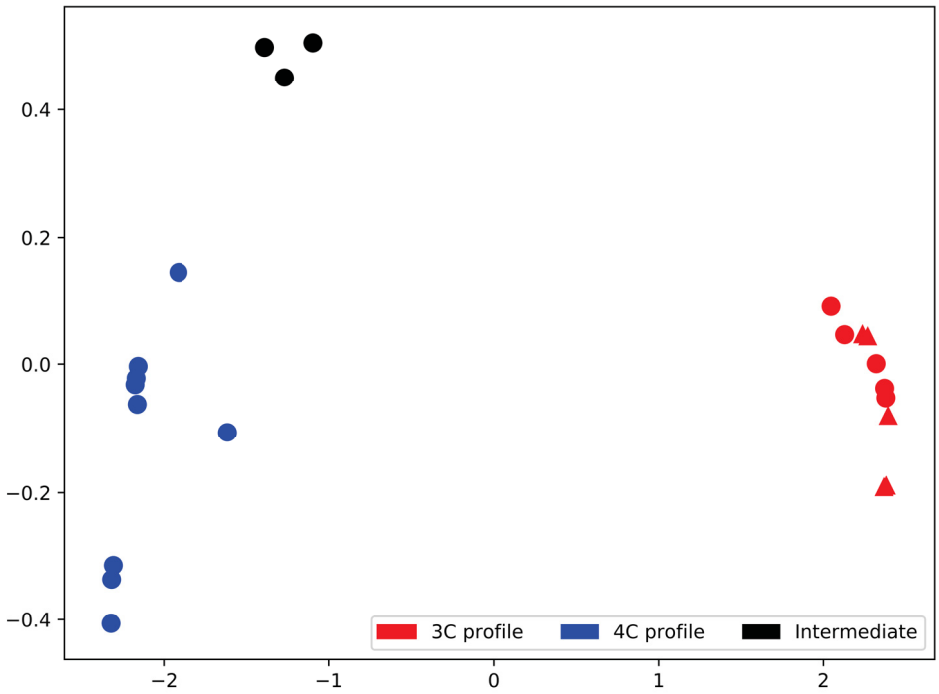



Figure 29: Multi dimensional scaling of the phenotypic distance between different metabolic phenotypes (Table 10). The phenotypic distance is measured in terms of euclidean distance. The MAM1 ecotypes are denoted with circles, while MAM2 ecotypes are represented by triangles.

Next, we shall analyse the metabolic phenotypes, i.e. the GSL profiles, corresponding to the 22 ecotypes from [52]. Details on the metabolic phenotypes are given in Table 10. An MDS plot scaling the phenotypic distances between different metabolic phenotypes $M_{P,i}(i = 1 \ldots 22)$ is shown in figure 29. We used Euclidean distance to calculate the phenotypic distance between the metabolic phenotypes. The euclidean distance $\Delta$ between phenotypes $M_{P,1}$ and $M_{P,2}$ is calculated by ($\Delta = \| M_{P,1} - M_{P,2} \|$). Thus, every metabolic phenotype $M_{P,i}$ gets a distance from other metabolic phenotypes. Similar to the previous analysis of the metabolic genotypes (cf. Fig. 28), also in Fig. 29 we see three separate clusters of ecotypes. We see a clear separation of the MAM2 ecotypes denoted by triangular markers from MAM1 ecotypes denoted by circular markers.

To gain some insights on how different genotypes and their associated metabolic phenotypes are linked, we analysed the genotypic and phenotypic distance matrices together. A summary of our analyses is shown in Figure 30, where the genotypic distances are plotted against the phenotypic distances. Every dot in the plot denotes a pair of ecotypes. For the purpose of visual convenience, the dots are colour coded based on their amino acid sequence similarity to MAM1 or MAM2 gene. While the red and blue dots represent pairs that constitute MAM1 and MAM2 ecotypes, respectively, green dots represent heterogeneous pairs. In the bottom left of the Fig. 30, we

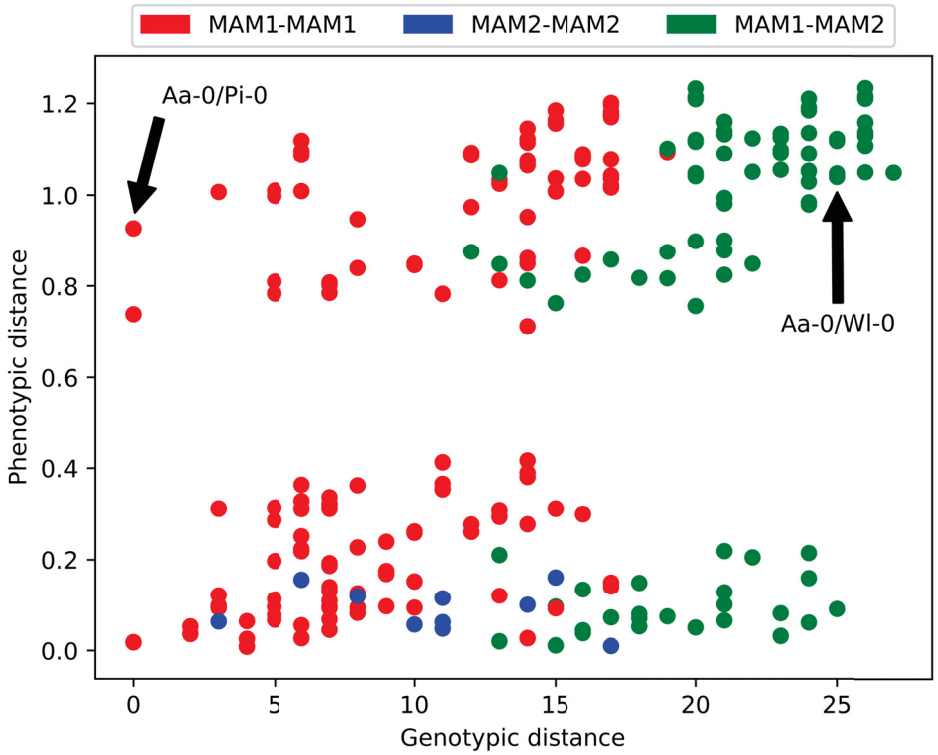



Figure 30: The genotypic versus phenotypic distance. Each dot represents a pair of ecotypes based on their sequence similarity to MAM1 or MAM2 genes. While the colours red and blue denote the pair, where both the ecotypes have MAM1 and MAM2 sequence, respectively, the green dots denote the heterogeneous pairs.

see ecotype pairs that are genetically similar and exhibit similar metabolic phenotypes. Whereas the top right of the Fig. 30 shows pairs (e.g. *Aa*−0/*Wl*−0), which are genetically very different and exhibit very different GSL profiles. The two aforestated cases are

understandable in the sense that similar metabolic genotypes shall exhibit similar metabolic phenotypes, and *vice-versa*. However, ecotype pairs located at the upper left (e.g. *Aa*—0/*Pi*—0) in Fig. 30 show high genetic similarity but high phenotypic variations. Investigating the factors affecting the variations in the metabolic phenotypes in this group of ecotypes will provide deeper insights into how genetic differences carve out the metabolic phenotypes. Moreover, information on the localisation of polymorphisms in GSL genes will provide a deeper understanding of how genetic differences are related to the enzymatic properties of the genes.

### 4.1.4 *Assessing the polymorphisms in GSL genes*

Polymorphisms in the active sites of an enzyme, in principle, can change the catalytic properties of the enzyme. A previous study showed that a mutagenesis of Serine to Phenylalanine, and Alanine to Threonine in MAM1 enzyme of *A. thaliana Columbia* ecotype lead to the loss of conversion of 3C to 4C glucosinolates [58]. However, the exact effect on the enzymatic properties is poorly understood. To gain some insights on the localisation of polymorphisms in the active sites of the metabolic enzymes, we extracted the information from the NCBI's conserved domain database [77]. The key MAM synthases, for example, are known to harbour their active sites in the region $92-294$ amino acid positions of the amino acid sequence. We used the information to assess the polymorphisms in the MAM synthases across the 22 Arabidopsis ecotypes.

Figure 31 shows a pairwise comparison of polymorphisms in the active site of MAM synthases to the total polymorphisms in the metabolic genotypes from 22 *A. thaliana* ecotypes. We clearly see that the number of substitutions in the active site of MAM synthases, increases with the the genotypic distance between the metabolic genotypes. Importantly, MAM synthases accumulate significantly

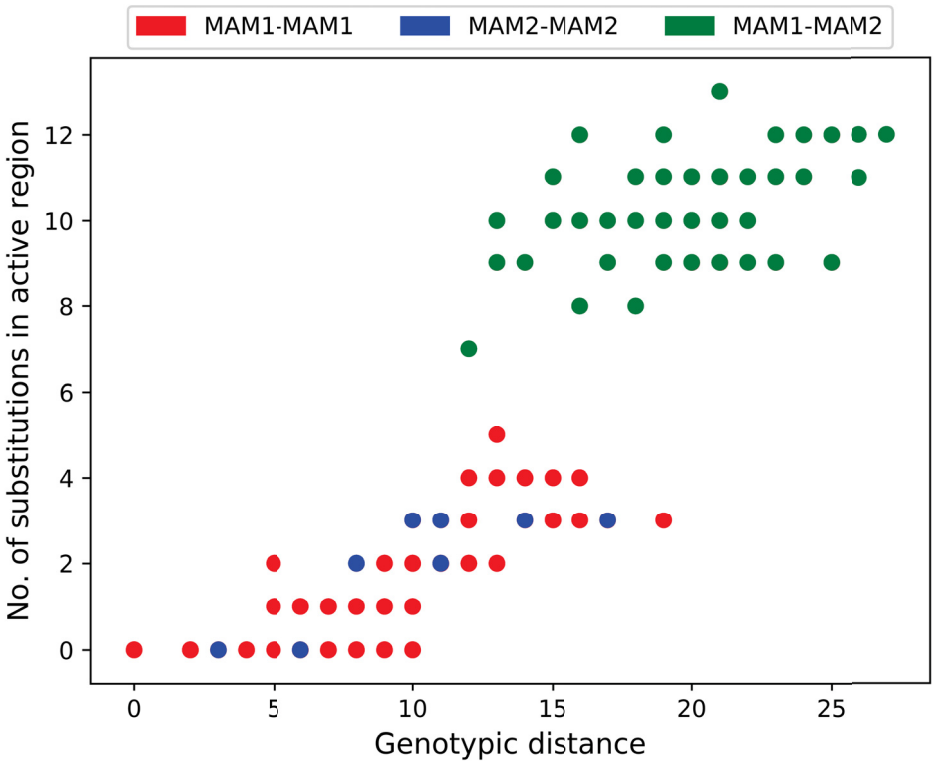



Figure 31: Polymorphisms in the active region of MAM synthases versus the genotypic distance between different metabolic genotypes. Each dot represents a pair of ecotypes. The red and blue dots denote pairs, where both the ecotypes possess MAM1 and MAM2 genes, respectively. Whereas, the green dots denote heterogeneous pairs.

higher number of substitutions in its active sites than any random substitution (Fishers test [25], $p < 0.001$).

## 4.2 DISCUSSION

Natural variation within species is regulated by complex networks of genes and their polymorphisms [26, 66]. Allelic composition at several glucosinolate biosynthetic loci results in different glucosinolate profiles among *Arabidopsis thaliana* ecotypes [52]. These variations, however, complicates our understanding of how genetic variations lead to metabolic properties of respective genes. The metabolic properties of a gene depend on the specific order of amino acids encoded by the DNA sequence of the respective gene. Thus, given a

similar environment for growth, enzymes with similar amino acid sequences are assumed to have similar intrinsic properties such as the dissociation constant ($K_d$) of enzyme-substrate (E-S) complexes in a reaction catalysed by the enzyme. However, the gene regulatory networks may change the maximum velocity of the enzyme-catalysed reaction. Moreover, polymorphisms in the active sites, where the reactions are catalysed, of an enzyme can change the $K_d$ of E-S complexes in a reaction catalysed by the enzyme. To dissect the behaviour originating from the genetic variations and gene regulatory networks, we investigate the level polymorphism in GSL biosynthesis genes from different Arabidopsis ecotypes.

The diversity analysis of GSL genes from 343 Arabidopsis ecotypes [1] showed that there exist genes that are highly diverse and highly conserved in their amino acid sequence across the 343 ecotypes (cf. Fig. 22). Our analyses showed a normal distribution of diversity in the genes active in three different stages of glucosinolate biosynthesis. The MAM synthases (MAM1/MAM2 and MAM3) showed a high level of polymorphism (high entropy) in their amino acid sequences, whereas BAT5 transporter showed low diversity (low entropy) across its amino acid sequence. A possible explanation of the aforementioned observation could be the high specificity of the MAM synthases towards respective substrates and high generality of BAT5 towards different substrates transported across the plastidial membrane. As BAT5 exerts activity on compounds that are a part of both GSL biosynthesis and Met-salvage pathway [30, 88], thus, mutations in the coding region of BAT5 may impair the functioning of both pathways. However, a further analysis of the diversity in the coding region of BAT5 showed a high level of polymorphism but they are not reflected at the amino acid level (cf. Fig. 23). The results cannot be explained directly, but it could be a consequence of purifying natural selection. The analysis provides a general insight to the diversity of GSL genes across an example set 343

Arabidopsis ecotypes. However, it leaves a gap in our knowledge of presence (absence) of different GSL genes acroos these ecotypes, given the highly debatable MAM gene conundrum [59, 95].

To bridge the gap in our knowledge of presence (absence) of MAM synthases across different Arabidopsis ecotypes, we performed a phylogenetic analysis based on the sequence similarity of MAM1 and MAM2 genes. The phylogenetic analyses yielded a mid-point rooted tree (Fig. 24) with two major branches based on the similarity and differences to MAM1 and MAM2 amino acid sequences. Thus, it is plausible to assume that ecotypes that show high similarity to MAM2 sequence, actually possess MAM2 genes rather than MAM1 gene. Such kind of analyses is crucial for dissecting the effects arising out of presence (absence) of MAM1 gene. Moreover, it encourages for studies involving reannotation of sequences in different Arabidopsis ecotypes as the presence of MAM1 cannot be assumed automatically.

The investigation of genotypic and phenotypic distances between different Arabidopsis ecotypes showed that there exist ecotypes that have identical metabolic genotype but exhibit high diversity in their associated metabolic phenotype, and *vice versa* (cf. Fig. 30). To dissect the behaviour originating from the genetic differences and relating it to the enzymatic properties, we investigated the polymorphisms in the active sites of metabolic enzymes. In this study, we analysed the polymorphisms in the active sites of MAM synthases. To our surprise, the probability of observing a substitution in the active sites of MAM synthases is significantly higher compared to a random substitution across the genes. Polymorphisms in the active sites of MAM synthases, in principle, can alter the dissociation constants of the enzyme-substrate complex in MAM-catalysed reactions. Moreover, based on the level of polymorphisms in active sites, the amount of change in the dissociation constants can be expected across different ecotypes. To test this speculation, studies involving

several targetted mutations are required. With the current advancement of biotechnology, we can expect some research in the proposed direction.

The current research is replete with how genetic variations lead to different phenotypes. However, very less is known about how these variations are or can be associated to the enzymatic properties. Moreover, the link between metabolic genotypes and the associated phenotypes is poorly understood. In this study, by providing new insights of integrating the genetic information, we push the envelope a bit further towards having a deeper understanding of how the variations in a metabolic genotype can carve exclusive metabolic phenotypes.

Part IV

# SUMMARY, OUTLOOK AND CONCLUSIONS

# 5

# SUMMARY AND OUTLOOK

The presence of glucosinolates in the model plant *Arabidopsis thaliana* has benefitted glucosinolate research tremendously from the 'omics databases, bioinformatic tools, natural variation and mutant collections that are available for *A. thaliana*. Although the genes involved in the biosynthesis of glucosinolates are well known by now, much more discovery awaits before we fully understand why and how plants synthesise certain glucosinolates with particular frequencies.

Till date, more than 135 structurally-different glucosinolates are known [2]. The structural diversity of glucosinolates arises from a three-phase biosynthesis. First, an aliphatic and aromatic amino acids are elongated by adding a methylene group into their side chains. Second, the amino acid moeity is metabolically reconfigured to form a core-structure of glucosinolates. Third, the formed glucosinolates are modified by different secondary modifications. The chain-elongation process has gained considerable attention over the past decades, as it is the branching point for the diversion of metabolic flux from primary to secondary metabolism.

In this study, we have developed a mathematical model of the chain-elongation of Met-derived glucosinolates in *A. thaliana*, based on our knowledge of the pathway structure and kinetic properties of the metabolic enzymes. The presented model is based on a similar idea conceived by Knoke et al. [57] concerning the chain-length distribution of Met-derived glucosinolates. We extend the model capabilities, by employing rate laws that account for the broad-range substrate specificity of metabolic enzymes. By providing the math-

ematical description, we illustrate how different biosynthetic rates are affected by all metabolite concentrations, a behaviour originating from broad-range substrate specificity of the metabolic enzymes. With our model, we intend to give a detailed mathematical representation of the underlying reactions catalysed by broad-range substrate specific enzymes, which is very important for fitting experimental data in the best possible way.

With the adjusted set of kinetic parameters, we were able to reproduce the GSL profiles of wild-type and mutant backgrounds of *A. thaliana* Columbia ecotype. Furthermore, we showcased how the model parameters are fitted to individually reproduce patterns of GSL accumulation observed in different *A. thaliana* ecotypes.

Significant progress has been made in the past decades in making the glucosinolate pathway a model system in ecology and evolution to identify causal genes underlying natural variation and testing consequences of polymorphisms at different biosynthetic loci. This, however, represents only the first dabble into the true complexity of this system. The metabolic diversity of glucosinolates, like any other metabolites, is carved by the genome (g) and the environment (e). Thus, a metabolic phenotype (P) is an interplay of the genome and environment,

$$g \times e \longrightarrow P.$$

The genome encodes amino acid sequences that define the catalytic properties of the enzyme, whereas the environment additionally influences the gene expression, which in turn regulates enzyme abundance and thus the maximum velocity of respective enzyme-catalysed reaction. However, it complicates our understanding of metabolic diversity as we often study only one ecotype. Thus, to fully understand the diversity of glucosinolates will require studies involving many ecotypes.

An important aim of mathematical models is to provide theoretical predictions that can be later verified by experiments. Our model provides a theoretical framework, where the link between the metabolic genotype and the associated phenotype can be investigated. With some preliminary results, we provide new insights into how different metabolic phenotypes can be produced by varying the enzymatic properties of metabolic enzymes.

Intuitively, one would assume that a genetically similar ecotype should exhibit a similar metabolic phenotype, given a similar conditions for growth. The dissociation constants of the enzyme-substrate (E-S) complexes in an enzyme-catalysed reaction depend on the coding region of respective enzyme, whereas the maximum velocity of the reaction depends on the promoter regions of DNA that initiates transcription of a particular gene. Thus, for genetically-similar ecotypes, it is plausible to assume similar values of dissociation constants of E-S complexes, while different maximum velocities can still exist. From our analyses (Fig.30), we know that there exist a few Arabidopsis ecotypes that have identical (zero genotypic distance) metabolic genotypes but very different metabolic phenotypes. We hypothesize that by varying the enzyme expressions, which eventually translates to enzyme concentration, in identical metabolic genotypes, different metabolic phenotypes can be produced. To investigate how such metabolic genotypes can exhibit different metabolic phenotypes, we selected an ecotype pair, $Aa-0/Pi-0$, which composes identical metabolic genotypes with highest phenotypic distance (cf. Fig. 30). By adjusting the model parameters, we were able to reproduce the chain-length distribution of GSLs in the wild-type $Aa-0$ ecotype. The kinetic parameters that reproduced the $Aa-0$ profile will be referred to as fiducial parameters, and are given in Table 6. Further, we used the fiducial parameters to vary the concentrations $E_t$ of different metabolic enzymes to reproduce the GSL profile of the wild-type $Pi-0$ ecotype. By varying $E_t$, one changes the $V_{max}$

of the enzyme-catalysed reaction as $V_{max} = E_t \cdot k_{cat}$, where $k_{cat}$ is the turnover number. Figure 32 shows a comparison between the model simulated and experimentally-observed GSL profiles from wild-type $Aa-0$ and $Pi-0$ ecotypes. Model simulations showed

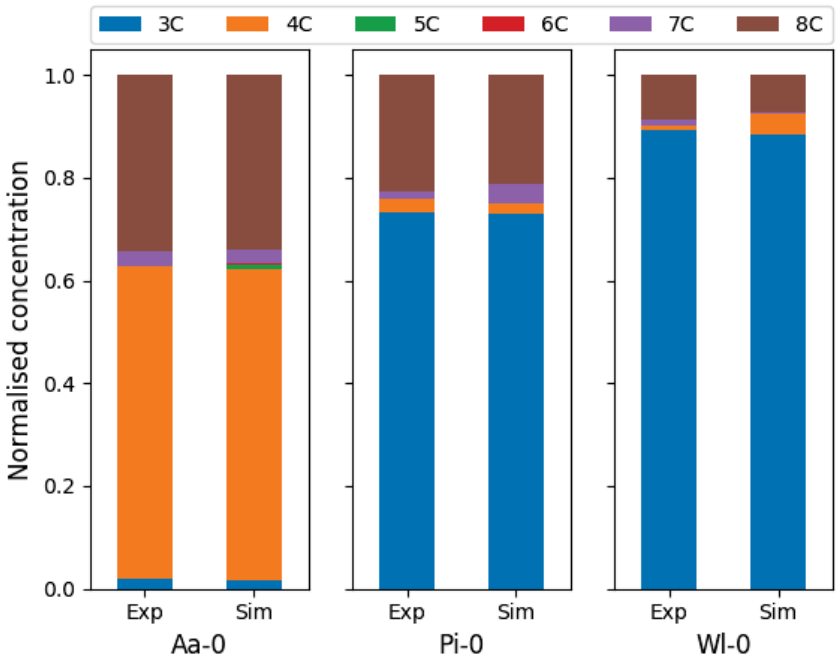



Figure 32: A comparison between the model simulated (Sim) and experimentally-measured (Exp) GSL profiles of wild-type $Aa-0$, $Pi-0$, and $Wl-0$ ecotypes. Length of the bars are normalised to the total accumulation of Met-derived GSLs in repective ecotypes. The kinetic parameters used for simulation are given in Table 6.

a good agreement with the experimental observation. Moreover, it showcases that by varying $V_{max}$ of different enzyme-catalysed reactions different metabolic phenotypes can be produced. Furthermore, to investigate whether or not the GSL profile of a genetically different ecotype can be reproduced by varying $V_{max}$, in terms of enzyme concentrations. For this purpose, we selected $Wl-0$ ecotype, which has very different metabolic genotype than $Aa-0$, and estimating the enzymatic properties is not intuitive. To our surprise, by varying the $V_{max}$ values in the fiducial parameters we managed to reproduce the GSL profile of $Wl-0$ ecotype. The agreement between the experimentally measured and model simulated GSL concentra-

tions is shown in Fig. 32. It, however, complicates our understanding of whether the exhibited metabolic phenotype originates from the regulatory or genetic plasticity. Moreover, dissecting the effect from one plasticity versus the other will require the analysis of metabolic phenotypes that could emerge due to all possible combinations of dissociation constants and maximum velocities.

Model simulations showed that by just changing the enzyme concentrations, in principle, one can reproduce different metabolic phenotypes. Similarly, the dissociation constants with respect to different substrates of respective enzyme-catalysed reactions can be altered to reproduce different metabolic phenotypes. The enzyme concentrations are regulated by the gene regulatory network, whereas the dissociation constants are regulated by the genome. While changes in gene regulation can take place in due course of plants life cycle due to fluctuating physical environment, the changes in genetics occur in evolutionary time-scale. It is not straightforward to understand why changes in evolutionary time-scale are required, when a similar outcome can be achieved by instant (faster) adjustments. It could be a consequence of neutral or random mutations that plants accumulate due to changing environmental conditions. A typical example of such evolutionary changes is the existence of highly variable MAM1 and MAM2 genes across different Arabidopsis ecotypes. Although the metabolic diversity created due to the presence of either of the MAM genes can be compensated by the regulatory plasticity, the coding regions of MAM1 and MAM2 evolved in due course of evolution. We can speculate it to be a strategy to increase plants fitness in challenging environmental conditions. Nevertheless, the faster adjustments, i.e. the alterations in $V_{max}$, could be a strategy adopted by plants to combat sudden demands of a specific metabolic phenotype.

To fully understand why plants produce different phenotypes, one would need to perform studies involving $n$ plant ecotypes under $m$ physiological conditions;

$$n \times m \longrightarrow Model \longrightarrow P.$$

Thus, by integrating the experimental data, our model can be used to predict the conditional probabilities of a particular phenotype. An interesting future strategy would be to employ the model to predict posterior plausibility based on the prior distribution of model parameters.

# 6

# CONCLUSIONS

It is fortunate for glucosinolate research that the glucosinolate-producing *Arabidopsis thaliana* was selected as the model plant to have its genome sequenced, a choice that was based on its small genome and short life cycle [71]. The glucosinolate research has benefitted tremendously from the availability of genetic sequences, mutant collections, and tools for expression profiling. In addition to the natural variation among Arabidopsis ecotypes, recombinant inbred lines and markers for mapping have greatly facilitated the elucidation of the biosynthetic pathway. Although much has been learned over past decades, much more of discovery awaits before we completely understand why and how plants synthesise glucosinolates. Application of systems biology approaches to link genetic, protein, and metabolite data should ensure further advancement in glucosinolate research. Precisely, identification of major regulatory factors that control flux through the biosynthetic pathway will allow the metabolic engineering of glucosinolate profiles to progress from the empirical to the predictive stage.

A particular difficulty in the analysis of glucosinolates is the vast diversity of chemical-structures. As much as it is true for experimental identification, theoretical descriptions are equally challenging. In this work, we focussed our attention on the most-abundant class of aliphatic glucosinolates, which are derived from methionine (Met), found in *Arabidopsis thaliana*. The structural diversity of methionine-derived glucosinolates arises from variation in the side chains. Given the broad-range substrate specificity of metabolic enzymes, it was impossible to draw conclusions on how changes in

kinetic properties would result the accumulation of glucosinolates by simple inspection. The corresponding metabolic network of the chain-elongation pathway is so complex that mathematical modelling is required to understand and investigate the regulators of metabolic flux.

In this study, we have developed a mathematical model of the chain-elongation of Met-derived glucosinolates in *A. thaliana*, based on our knowledge of the pathway structure and kinetic properties of the metabolic enzymes. The model elucidates how Met-derived glucosinolates with a particular frequency are produced in *A. thaliana* plants.

With the adjusted set of kinetic parameters, model simulations for wild-type and knockout mutants showed a surprising agreement with the actual patterns of glucosinolate concentrations in *A. thaliana* Columbia ecotype. Model simulations allowed us to elucidate how different patterns of glucosinolate concentrations originate from knockdown rather than a knockout of metabolic enzymes, which cannot be assumed automatically.

A long-standing question in the evolution of glucosinolates is how certain plants can exhibit different metabolic phenotypes. By providing new insights to integrate the information on genetic differences and enzyme expressions in governing the enzymatic properties of metabolic enzymes, our model provides a framework where the link between the genotype and phenotype can be investigated.

The presented model is more powerful and its capabilities are far from being exploited. The Model will allow the posterior plausibilty based on the prior distributio of model parameters. Often, these parameters are not known or estimating these involves a tedious and costly task owing to extensive experimentation. In such cases, data-driven methods like artificial intelligence (AI) can be used to develop modelling applications, which can be used for parameter estimation. By incorporating the data-driven modelling formalisms

into the phenomenological models, we can bridge the gap in your knowledge of system parameters.

On the final remarks, we would like to mention that the principal aim of this work was to develop a mathematical model of glucosinolate metabolism, and elucidate how plants accumulate certain glucosinolates with particular frequency. By providing insights into metabolic diversity, we provide a theoretical framework to investigate how exclusive traits emerge out of genomic variations. Thus, we push the boundaries of our understanding of how the interplay of the genome and enviornment shapes metabolic phenotypes.

# Part V

# APPENDIX

# A

# DIRECTIONS TO USE THE MODEL

The calculations presented in this study have been facilitated by computer software developed as a part of this research project. The software architecture is designed using the programming language Python. The software is available on Gitlab (www.gitlab.com), and can be accessed by request. A detailed documentation is being written presently, and it will be available for public use soon.

## A.1 SOFTWARE ARCHITECTURE

In order to be able to describe the model capabilities, we provide an overview of the software architecture. This documentation is an effort to facilitate other programmers to use and extend or modify the existing packages. An overview of the model structure is depicted in Fig. 33.

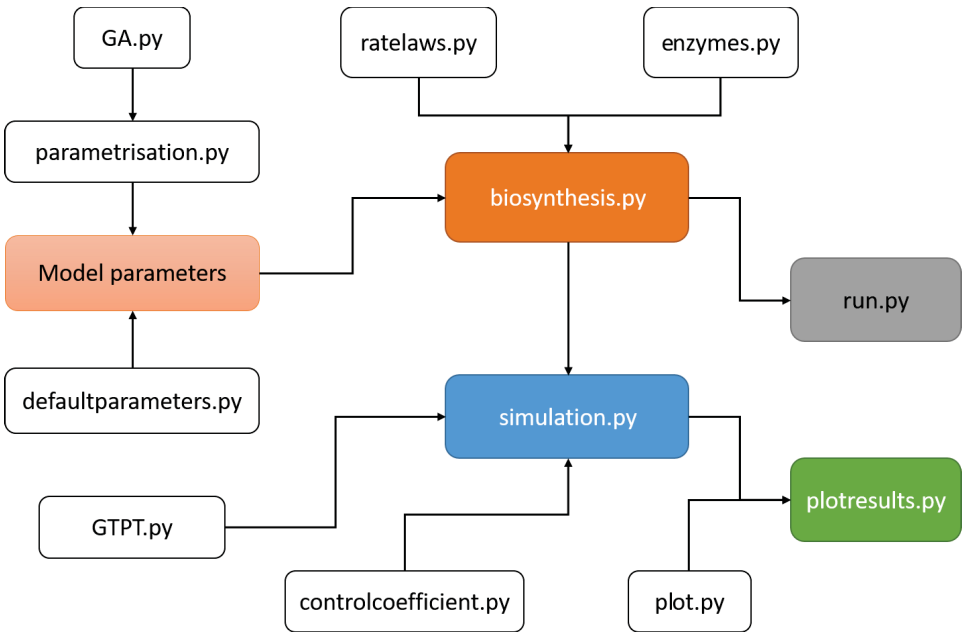



Figure 33: A overview of the software architecture

*biosynthesis.py*

Instance of this class is used to describe the biosynthesis of metabolic intermediates ($A_i$ and $K_i$). It uses methods defining rate laws and kinetic parameters. By keeping the attributes free, we allow flexibility and a wide range of implementation of the biosynthesis model.

*run.py*

The method defined in this python file can be used to simulate a typical steady-state concentration of metabolic intermediates of the biosynthesis model with a reference parameters.

*simulation.py*

Instance of this class is used to perform the analyses presented in this study. By keeping the attributes free, we provide high flexibility in the implementation of different methods.

*plotresults.py*

It exclusively composes the methods used to produce plots presented in this study.

# B

## DERIVATION OF KINETIC RATE LAW FOR MONOMOLECULAR REACTIONS

To derive enzymatic rate laws, usually certain assumptions are employed. For example, the rapid-equilibrium assumption says that the enzyme bound to intermediates are at equilibrium with the substrates, or the quasi-steady-state assumption, which says that the enzymic intermediates attain a quasi-steady-state even when the concentrations of the nonenzymic reactants still change in time. In principle, the steady state rate equations can be derived in the same way as two-step Michaelis-Menten mechanism [72]. First leg starts by writing down the expressions for rate of change of concentrations of all reactions catalysed by one enzyme (isoform). Secondly, by setting these expression to zero, write an equation to express the sum of all the concentrations as constant. Lastly, solve the set of ordinary differential equations.

The reaction scheme for $n$ substrates competing for the binding site of an enzyme $E$:

$$\begin{aligned} S_1 + E &\underset{k_{-1,1}}{\overset{k_{1,1}}{\rightleftarrows}} ES_1 \underset{k_{-2,1}}{\overset{k_{2,1}}{\rightleftarrows}} P_1 + E \\ S_2 + E &\underset{k_{-1,2}}{\overset{k_{1,2}}{\rightleftarrows}} ES_2 \underset{k_{-2,2}}{\overset{k_{2,2}}{\rightleftarrows}} P_2 + E \\ &\vdots \\ S_n + E &\underset{k_{-1,n}}{\overset{k_{1,n}}{\rightleftarrows}} ES_6 \underset{k_{-2,n}}{\overset{k_{2,n}}{\rightleftarrows}} P_n + E \end{aligned} \tag{B.1}$$

The list of differential equations for the metabolites in Scheme B.1:

$$\frac{dS_1}{dt} = -k_{1,1}E \cdot S_1 + k_{-1,1}ES_1$$
$$\frac{dS_2}{dt} = -k_{1,2}E \cdot S_2 + k_{-1,2}ES_2$$
$$\vdots$$
$$\frac{dS_n}{dt} = -k_{1,n}E \cdot S_6 + k_{-1,n}ES_6$$

$$\frac{dE}{dt} = -(k_{1,1}E \cdot S_1 + k_{1,2}E \cdot S_2 + \ldots + k_{1,n}E \cdot S_n)$$
$$+(k_{-1,1}ES_1 + k_{-1,2}ES_2 + \ldots + k_{-1,n}ES_n)$$
$$+(k_{1,1}E \cdot P_1 + k_{1,2}E \cdot P_2 + \ldots + k_{1,n}E \cdot P_n)$$

$$\frac{dES_1}{dt} = k_{1,1}E \cdot S_1 - k_{-1,1}ES_1 - k_{2,1}ES_1 + k_{-2,1}E \cdot P_1$$
$$\frac{dES_2}{dt} = k_{1,2}E \cdot S_1 - k_{-1,2}ES_1 - k_{2,2}ES_1 + k_{-2,2}E \cdot P_2$$
$$\vdots$$
$$\frac{dES_n}{dt} = k_{1,n}E \cdot S_1 - k_{-1,n}ES_1 - k_{2,n}ES_1 + k_{-2,n}E \cdot P_n$$
$$\frac{dP_1}{dt} = k_{2,1}ES_1 - k_{-2,n}E \cdot P_n$$
$$\frac{dP_2}{dt} = k_{2,2}ES_2 - k_{-2,n}E \cdot P_n$$
$$\vdots$$
$$\frac{dP_n}{dt} = k_{2,n}ES_n - k_{-2,n}E \cdot P_n$$

For n = 6;

Conservation relation:

$$E_t = E + ES_1 + ES_2 + ES_3 + ES_4 + ES_5 + ES_6 \tag{B.2}$$

$\frac{dES_1}{dt} = 0$ :

$$k_{1,1}E \cdot S_1 - k_{-1,1}ES_1 - k_{2,1}ES_1 + k_{-2,1}E \cdot P_1 = 0$$

$$E = \frac{(k_{-1,1} + k_{2,1})ES_1}{k_{1,1}S_1 + k_{-2,1}P_1} \tag{B.3}$$

$\frac{dES_2}{dt} = 0$ :

$$E = \frac{(k_{-1,2} + k_{2,2})ES_1}{k_{1,2}S_2 + k_{-2,2}P_2}. \tag{B.4}$$

Since, Eq. B.3 = Eq. B.4:

$$ES_2 = ES_1 \frac{(k_{-1,1} + k_{2,1})(k_{1,2}S_2 + k_{-2,2}P_2)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,2} + k_{2,2})} \tag{B.5}$$

$\frac{dES_3}{dt} = 0$ :

$$E = \frac{(k_{-1,3} + k_{2,3})}{k_{1,3}S_3 + k_{-2,3}P_3} \tag{B.6}$$

$\frac{dES_4}{dt} = 0$ :

$$E = \frac{(k_{-1,4} + k_{2,4})}{k_{1,4}S_4 + k_{-2,4}P_4} \tag{B.7}$$

$\frac{dES_5}{dt} = 0$ :

$$E = \frac{(k_{-1,5} + k_{2,5})}{k_{1,5}S_5 + k_{-2,5}P_5} \tag{B.8}$$

$\frac{dES_6}{dt} = 0$ :

$$E = \frac{(k_{-1,6} + k_{2,6})}{k_{1,6}S_6 + k_{-2,6}P_6} \tag{B.9}$$

Using Eq. B.6, B.7, B.8 and B.9:

$$ES_3 = ES_1 \frac{(k_{-1,1} + k_{2,1})(k_{1,3}S_3 + k_{-2,3}P_3)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,3} + k_{2,3})}$$
$$ES_4 = ES_1 \frac{(k_{-1,1} + k_{2,1})(k_{1,4}S_4 + k_{-2,4}P_4)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,4} + k_{2,4})}$$
$$ES_5 = ES_1 \frac{(k_{-1,1} + k_{2,1})(k_{1,5}S_5 + k_{-2,5}P_5)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,5} + k_{2,5})}$$
$$ES_6 = ES_1 \frac{(k_{-1,1} + k_{2,1})(k_{1,6}S_6 + k_{-2,6}P_6)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,6} + k_{2,6})}$$

From Eq. B.2:

$$E_t = ES_1 \Bigg( 1 + \frac{(k_{-1,1} + k_{2,1})ES_1}{k_{1,1}S_1 + k_{-2,1}P_1} + \frac{(k_{-1,1} + k_{2,1})(k_{1,2}S_2 + k_{-2,2}P_2)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,2} + k_{2,2})} + \frac{(k_{-1,1} + k_{2,1})(k_{1,3}S_3 + k_{-2,3}P_3)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,3} + k_{2,3})} + \frac{(k_{-1,1} + k_{2,1})(k_{1,4}S_4 + k_{-2,4}P_4)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,4} + k_{2,4})} + \frac{(k_{-1,1} + k_{2,1})(k_{1,5}S_5 + k_{-2,5}P_5)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,5} + k_{2,5})} + \frac{(k_{-1,1} + k_{2,1})(k_{1,6}S_6 + k_{-2,6}P_6)}{(k_{1,1}S_1 + k_{-2,1}P_1)(k_{-1,6} + k_{2,6})} \Bigg)$$

$$ES_1 = \frac{E_t \frac{(k_{-1,1}+k_{2,1})ES_1}{k_{1,1}S_1+k_{-2,1}P_1}}{(1 + \frac{(k_{1,2}S_2+k_{-2,2}P_2)}{(k_{-1,2}+k_{2,2})} + \ldots + \frac{(k_{1,6}S_6+k_{-2,6}P_6)}{(k_{-1,6}+k_{2,6})}} \tag{B.10}$$

$$E = \frac{E_t}{(1 + \frac{(k_{1,2}S_2+k_{-2,2}P_2)}{(k_{-1,2}+k_{2,2})} + \ldots + \frac{(k_{1,6}S_6+k_{-2,6}P_6)}{(k_{-1,6}+k_{2,6})}} \tag{B.11}$$

Using Eq. B.10 and B.11 $\rightarrow \frac{dP_1}{dt}$:

$$\frac{dP_1}{dt} = k_{2,1}ES_1 - k_{-2,1}E \cdot P_1$$

$$\frac{dP_1}{dt} = \frac{k_{2,1}E_t\frac{S_1}{K^+_{d,1}} - k_{-1,1}E_t\frac{P_1}{K^-_{d,1}}}{\frac{S_1}{K^+_{d,1}} + \frac{S_2}{K^+_{d,2}} + \ldots + \frac{S_6}{K^+_{d,6}} + \frac{P_1}{K^-_{d,1}} + \frac{P_2}{K^-_{d,2}} + \ldots + \frac{P_6}{K^-_{d,6}} + 1}$$
$$= \frac{k_{2,1}E_t\frac{S_1}{K^+_{d,1}} - k_{-1,1}E_t\frac{P_1}{K^-_{d,1}}}{\sum_{i=1}^{6}\left(\frac{S_i}{K^+_{d,i}} + \frac{P_i}{K^-_{d,i}}\right) + 1}$$

Thus, the generalised rate law for the production of product $P_i$ with n competing substrates for the binding of an enzyme *E* reads:

$$\frac{dP_i}{dt} = \frac{V^+_m\frac{S_i}{K^+_{d,i}} - V^-_m\frac{P_i}{K^-_{d,i}}}{\sum_{j=1}^{n}\left(\frac{S_j}{K^+_{d,j}} + \frac{P_j}{K^-_{d,j}}\right) + 1}, \tag{B.12}$$

where $V^+_m = k_{2,i}E_t$ and $V^-_m = k_{-1,i}E_t$.

# C

# REACTIONS AND KINETIC PARAMETERS

## C.1 LIST OF REACTIONS USED FOR MODELLING

Table 3: List of reactions used for modelling the chain-elongation of Met-derived glucosinolates.

| Symbol | Reaction | Enzymes |
|---|---|---|
| R1 | $K_{p,0} \xrightarrow{MAM_3} K_{p,1}$ | MAM3 |
| R2 | $K_{p,1} \xrightarrow{MAM_3} K_{p,2}$ | |
| R3 | $K_{p,2} \xrightarrow{MAM_3} K_{p,3}$ | |
| R4 | $K_{p,3} \xrightarrow{MAM_3} K_{p,4}$ | |
| R5 | $K_{p,4} \xrightarrow{MAM_3} K_{p,5}$ | |
| R6 | $K_{p,5} \xrightarrow{MAM_3} K_{p,6}$ | |
| R7 | $K_{p,0} \xrightarrow{MAM_1} K_{p,1}$ | MAM1 |
| R8 | $K_{p,1} \xrightarrow{MAM_1} K_{p,2}$ | |
| R9 | $K_{c,0} \xrightarrow{BAT_5} K_{p,0}$ | BAT5 |
| R10 | $K_{c,1} \xrightarrow{BAT_5} K_{p,1}$ | |
| R11 | $K_{c,2} \xrightarrow{BAT_5} K_{p,2}$ | |
| R12 | $K_{c,3} \xrightarrow{BAT_5} K_{p,3}$ | |
| R13 | $K_{c,4} \xrightarrow{BAT_5} K_{p,4}$ | |
| R14 | $K_{c,5} \xrightarrow{BAT_5} K_{p,5}$ | |
| R15 | $K_{c,6} \xrightarrow{BAT_5} K_{p,6}$ | |
| R16 | $K_{p,0} \xrightarrow{BAT_5} K_{c,0}$ | |
| R17 | $K_{p,1} \xrightarrow{BAT_5} K_{c,1}$ | |
| R18 | $K_{p,2} \xrightarrow{BAT_5} K_{c,2}$ | |
| R19 | $K_{p,3} \xrightarrow{BAT_5} K_{c,3}$ | |

*continues on next page*

| | | |
|---|---|---|
| R20 | $K_{p,4} \xrightarrow{BAT_5} K_{c,4}$ | |
| R21 | $K_{p,5} \xrightarrow{BAT_5} K_{c,5}$ | |
| R22 | $K_{p,6} \xrightarrow{BAT_5} K_{c,6}$ | |
| R23 | $A_{c,0} \xrightarrow{AATR} A_{p,0}$ | AATR |
| R24 | $A_{c,1} \xrightarrow{AATR} A_{p,1}$ | |
| R25 | $A_{c,2} \xrightarrow{AATR} A_{p,2}$ | |
| R26 | $A_{c,3} \xrightarrow{AATR} A_{p,3}$ | |
| R27 | $A_{c,4} \xrightarrow{AATR} A_{p,4}$ | |
| R28 | $A_{c,5} \xrightarrow{AATR} A_{p,5}$ | |
| R29 | $A_{c,6} \xrightarrow{AATR} A_{p,6}$ | |
| R30 | $A_{p,0} \xrightarrow{AATR} A_{c,0}$ | |
| R31 | $A_{p,1} \xrightarrow{AATR} A_{c,1}$ | |
| R32 | $A_{p,2} \xrightarrow{AATR} A_{c,2}$ | |
| R33 | $A_{p,3} \xrightarrow{AATR} A_{c,3}$ | |
| R34 | $A_{p,4} \xrightarrow{AATR} A_{c,4}$ | |
| R35 | $A_{p,5} \xrightarrow{AATR} A_{c,5}$ | |
| R36 | $A_{p,6} \xrightarrow{AATR} A_{c,6}$ | |
| R37 | $A_{c,0} \xrightarrow{BCAT_3} K_{c,0}$ | BCAT3 |
| R38 | $A_{c,1} \xrightarrow{BCAT_3} K_{c,1}$ | |
| R39 | $A_{c,2} \xrightarrow{BCAT_3} K_{c,2}$ | |
| R40 | $A_{c,3} \xrightarrow{BCAT_3} K_{c,3}$ | |
| R41 | $A_{c,4} \xrightarrow{BCAT_3} K_{c,4}$ | |
| R42 | $A_{c,5} \xrightarrow{BCAT_3} K_{c,5}$ | |
| R43 | $A_{c,6} \xrightarrow{BCAT_3} K_{c,6}$ | |
| R44 | $K_{c,0} \xrightarrow{BCAT_3} A_{c,0}$ | |
| R45 | $K_{c,1} \xrightarrow{BCAT_3} A_{c,1}$ | |
| R46 | $K_{c,2} \xrightarrow{BCAT_3} A_{c,2}$ | |
| R47 | $K_{c,3} \xrightarrow{BCAT_3} A_{c,3}$ | |
| R48 | $K_{c,4} \xrightarrow{BCAT_3} A_{c,4}$ | |
| R49 | $K_{c,5} \xrightarrow{BCAT_3} A_{c,5}$ | |
| R50 | $K_{c,6} \xrightarrow{BCAT_3} A_{c,6}$ | |
| R51 | $A_{c,0} \xrightarrow{BCAT_4} K_{c,0}$ | BCAT4 |
| R52 | $A_{c,1} \xrightarrow{BCAT_4} K_{c,1}$ | |

*continues on next page*

| | | |
|---|---|---|
| R53 | $A_{c,2} \xrightarrow{BCAT_4} K_{c,2}$ | |
| R54 | $A_{c,3} \xrightarrow{BCAT_4} K_{c,3}$ | |
| R55 | $A_{c,4} \xrightarrow{BCAT_4} K_{c,4}$ | |
| R56 | $A_{c,5} \xrightarrow{BCAT_4} K_{c,5}$ | |
| R57 | $A_{c,6} \xrightarrow{BCAT_4} K_{c,6}$ | |
| R58 | $K_{c,0} \xrightarrow{BCAT_4} A_{c,0}$ | |
| R59 | $K_{c,1} \xrightarrow{BCAT_4} A_{c,1}$ | |
| R60 | $K_{c,2} \xrightarrow{BCAT_4} A_{c,2}$ | |
| R61 | $K_{c,3} \xrightarrow{BCAT_4} A_{c,3}$ | |
| R62 | $K_{c,4} \xrightarrow{BCAT_4} A_{c,4}$ | |
| R63 | $K_{c,5} \xrightarrow{BCAT_4} A_{c,5}$ | |
| R64 | $K_{c,6} \xrightarrow{BCAT_4} A_{c,6}$ | |
| R65 | $A_{c,1} \xrightarrow{CYP_{79}F_1} A_{c,1}$ | CYP79F1 |
| R66 | $A_{c,2} \xrightarrow{CYP_{79}F_1} A_{c,2}$ | |
| R67 | $A_{c,3} \xrightarrow{CYP_{79}F_1} A_{c,3}$ | |
| R68 | $A_{c,4} \xrightarrow{CYP_{79}F_1} A_{c,4}$ | |
| R69 | $A_{c,5} \xrightarrow{CYP_{79}F_1} A_{c,5}$ | |
| R70 | $A_{c,6} \xrightarrow{CYP_{79}F_1} A_{c,6}$ | |
| R71 | $A_{c,5} \xrightarrow{CYP_{79}F_2} A_{c,5}$ | CYP79F2 |
| R72 | $A_{c,6} \xrightarrow{CYP_{79}F_2} A_{c,6}$ | |

Table 4: Fiducial set of kinetic parameters used for simulations. While the data marked with * is sourced from the works of [10, 55, 92, 97], the others are obtained from numerical optimisation. $p_V$, *vanilla* parameters; $p_C$, *Col* parameters; $p_1$, *Cvi* parameters

| Reaction | $K_m(\mu M)$ | | | $V'_{max}(\mu mol/min \cdot l)$ | | |
|---|---|---|---|---|---|---|
| | $p_V$ | $p_C$ | $p_1$ | $p_V$ | $p_C$ | $p_1$ |
| R1 | 1 | 932* | 810 | 1 | 2.896 | 1.163 |
| R2 | 1 | 476* | 80 | 1 | 2.99 | 0.705 |
| R3 | 1 | 463* | 221.741 | 1 | 5.738 | 2.901 |
| R4 | 1 | 300 | 836.548 | 1 | 3.4 | 13.003 |
| R5 | 1 | 253* | 69.226 | 1 | 0.728 | 21.438 |
| R6 | 1 | 81* | 1.4 | 1 | 0.062 | 13.003 |
| R7 | 1 | 3000* | 1568.875 | 1 | 47.61 | 124.334 |
| R8 | 1 | 640* | 707.495 | 1 | 75.737 | 344.747 |
| R9 | 1 | 1978.24 | 753.705 | 1 | 81.874 | 340.376 |
| R10 | 1 | 2138.53 | 829.396 | 1 | 88.47 | 315.179 |
| R11 | 1 | 2310.82 | 936.782 | 1 | 95.057 | 274.68 |
| R12 | 1 | 2482.88 | 1066.513 | 1 | 101.066 | 228.855 |
| R13 | 1 | 2639.84 | 1195.439 | 1 | 105.904 | 186.257 |
| R14 | 1 | 2766.2 | 1291.846 | 1 | 109.048 | 150.997 |
| R15 | 1 | 2848.3 | 3424.209 | 1 | 37.466 | 99.584 |
| R16 | 1 | 863.7 | 3729.357 | 1 | 39.913 | 98.21 |
| R17 | 1 | 933.68 | 4104.308 | 1 | 43.921 | 94.202 |
| R18 | 1 | 1008.91 | 4522.942 | 1 | 49.608 | 87.883 |
| R19 | 1 | 1084.031 | 4940.483 | 1 | 56.478 | 79.741 |
| R20 | 1 | 1152.56 | 5299.883 | 1 | 63.305 | 70.371 |
| R21 | 1 | 1207.728 | 5544.076 | 1 | 68.411 | 60.401 |
| R22 | 1 | 1243.576 | 1467.988 | 1 | 225.739 | 232.288 |
| R23 | 1 | 2474.039 | 1563.868 | 1 | 240.755 | 234.203 |
| R24 | 1 | 2638.601 | 1720.92 | 1 | 255.758 | 224.668 |
| R25 | 1 | 2803.037 | 1943.737 | 1 | 269.858 | 204.497 |
| R26 | 1 | 2957.571 | 2212.916 | 1 | 282.111 | 178.026 |
| R27 | 1 | 3091.854 | 2480.426 | 1 | 291.621 | 150.36 |
| R28 | 1 | 3196.086 | 2680.462 | 1 | 297.653 | 124.929 |

*continues on next page*

| | | | | | | |
|---|---|---|---|---|---|---|
| R29 | 1 | 3262.19 | 2448.426 | 1 | 10.323 | 61.685 |
| R30 | 1 | 410.615 | 2399.944 | 1 | 8.301 | 66.684 |
| R31 | 1 | 437.434 | 2260.182 | 1 | 41.173 | 72.056 |
| R32 | 1 | 481.364 | 2045.097 | 1 | 8.345 | 77.421 |
| R33 | 1 | 543.689 | 1777.922 | 1 | 1.499 | 82.315 |
| R34 | 1 | 618.982 | 1485.046 | 1 | 0.222 | 86.256 |
| R35 | 1 | 693.808 | 1191.777 | 1 | 3.453 | 88.816 |
| R36 | 1 | 749.76 | 534.549 | 1 | 27.26 | 282.541 |
| R37 | 1 | 2000 | 582.185 | 1 | 23.695 | 274.891 |
| R38 | 1 | 2000 | 640.718 | 1 | 27.001 | 254.183 |
| R39 | 1 | 2000 | 706.071 | 1 | 27.656 | 224.246 |
| R40 | 1 | 2000 | 771.252 | 1 | 5.281 | 189.757 |
| R41 | 1 | 2000 | 827.358 | 1 | 0.818 | 154.822 |
| R42 | 1 | 2000 | 865.479 | 1 | 13.129 | 122.301 |
| R43 | 1 | 2000 | 459.268 | 1 | 0.036 | 77.744 |
| R44 | 1 | 1000 | 405.302 | 1 | 0.036 | 77.744 |
| R45 | 1 | 1000 | 278.56 | 1 | 0.037 | 77.744 |
| R46 | 1 | 1000 | 149.102 | 1 | 0.04 | 77.77 |
| R47 | 1 | 1000 | 62.155 | 1 | 0.047 | 78.608 |
| R48 | 1 | 1000 | 20.179 | 1 | 0.057 | 88.265 |
| R49 | 1 | 1000 | 5.102 | 1 | 0.067 | 124.898 |
| R50 | 1 | 1000 | 2586.622 | 1 | 0.002 | 13.804 |
| R51 | 1 | 45* | 2758.673 | 1 | 0.002 | 11.422 |
| R52 | 1 | 45 | 2930.591 | 1 | 0.002 | 7.39 |
| R53 | 1 | 45 | 3092.157 | 1 | 0.002 | 3.75 |
| R54 | 1 | 45 | 3232.552 | 1 | 0.002 | 1.511 |
| R55 | 1 | 45 | 3341.527 | 1 | 0.003 | 0.533 |
| R56 | 1 | 45 | 3410.639 | 1 | 0.003 | 0.187 |
| R57 | 1 | 45 | 0 | 1 | 6.256 | 0.262 |
| R58 | 1 | 930* | 3.041 | 1 | 10.07 | 0.289 |
| R59 | 1 | 930 | 35.748 | 1 | 0.138 | 0.318 |
| R60 | 1 | 930 | 209.827 | 1 | 0.15 | 0.347 |
| R61 | 1 | 930 | 200.748 | 1 | 5.684 | 0.373 |
| R62 | 1 | 930 | 6.481 | 1 | 3.604 | 0.39 |
| R63 | 1 | 930 | 457.865 | 1 | 49.377 | 1.265 |
| R64 | 1 | 930 | 321.192 | 1 | 69.87 | 0.767 |
| R65 | 1 | 30 | 0 | 1 | 12.7 | 2.901 |

*continues on next page*

| | | | | | | |
|---|---|---|---|---|---|---|
| R66 | 1 | 34* | 9.899 | 1 | 42.5 | 0.238 |
| R67 | 1 | 37* | 2586.638 | 1 | 33.067 | 94.295 |
| R68 | 1 | 194* | 2550.961 | 1 | 35.746 | 100.567 |
| R69 | 1 | 216* | 2446.855 | 1 | 38.626 | 106.834 |
| R70 | 1 | 74* | 2282.7 | 1 | 41.502 | 112.724 |
| R71 | 1 | 374* | 2071.218 | 1 | 44.126 | 117.842 |
| R72 | 1 | 26* | 1827.843 | 1 | 46.238 | 121.815 |

Table 5: Kinetic parameters used for simulating the glucosinolate profiles of *Aa*−0, *Pi*−0 and *Wl*−0 Arabidopsis ecotypes. $p_3$, *Aa*−0 parameters; $p_4$, *Pi*−0 parameters; $p_5$, *Wl*−0 parameters.

| Reaction | $K_m(\mu M)$ | | | $V'_{max}(\mu mol/min \cdot l)$ | | |
|---|---|---|---|---|---|---|
| | $p_3$ | $p_4$ | $p_5$ | $p_3$ | $p_4$ | $p_5$ |
| R1 | 932 | 1286.23 | 103.47 | 1.73 | 21.2 | 12.84 |
| R2 | 476 | 276.04 | 22.21 | 1.79 | 30 | 7.79 |
| R3 | 463 | 2358.97 | 2563.41 | 3.44 | 4.25 | 2.41 |
| R4 | 300 | 2081.79 | 2262.2 | 2.04 | 19.04 | 10.82 |
| R5 | 253 | 1430.79 | 1554.78 | 0.44 | 31.4 | 17.84 |
| R6 | 81 | 765.85 | 832.22 | 0.04 | 19.04 | 10.82 |
| R7 | 3000 | 1184.75 | 4296.72 | 1.31 | 29.4 | 74.7 |
| R8 | 640 | 32.22 | 1374.07 | 2.08 | 13.18 | 109.42 |
| R9 | 1978.24 | 34.32 | 1463.81 | 2.25 | 11.54 | 119.31 |
| R10 | 2138.54 | 37.77 | 1610.82 | 2.43 | 8.52 | 126.75 |
| R11 | 2310.82 | 42.66 | 1819.38 | 2.61 | 5.33 | 130.49 |
| R12 | 2482.88 | 48.57 | 2071.34 | 2.78 | 2.87 | 130.88 |
| R13 | 2639.84 | 54.44 | 2321.73 | 2.91 | 1.37 | 129.48 |
| R14 | 2766.2 | 58.83 | 2508.97 | 3 | 0.59 | 127.93 |
| R15 | 2848.31 | 196.45 | 572.37 | 76.25 | 2.99 | 23.36 |
| R16 | 863.7 | 196.45 | 577.47 | 81.23 | 2.83 | 22.1 |
| R17 | 933.69 | 196.45 | 596.21 | 89.38 | 2.4 | 18.71 |
| R18 | 1008.91 | 196.51 | 648.41 | 100.96 | 1.82 | 14.17 |
| R19 | 1084.03 | 198.63 | 756.54 | 114.94 | 1.23 | 9.61 |
| R20 | 1152.56 | 223.03 | 917.52 | 128.83 | 0.75 | 5.83 |
| R21 | 1207.73 | 315.6 | 1075.13 | 139.22 | 0.41 | 3.16 |
| R22 | 1243.58 | 2382.82 | 139.69 | 459.4 | 0.25 | 95.73 |
| R23 | 2474.04 | 2254.05 | 140.93 | 489.96 | 0.25 | 90.56 |

*continues on next page*

| | | | | | | |
|---|---|---|---|---|---|---|
| R24 | 2638.6 | 1908.01 | 145.51 | 520.5 | 0.25 | 76.66 |
| R25 | 2803.04 | 1445.25 | 158.25 | 549.19 | 0.25 | 58.07 |
| R26 | 2957.57 | 979.61 | 184.64 | 574.13 | 0.25 | 39.36 |
| R27 | 3091.85 | 594.16 | 223.93 | 593.48 | 0.28 | 23.87 |
| R28 | 3196.09 | 322.48 | 262.39 | 605.76 | 0.4 | 12.96 |
| R29 | 3262.19 | 2184.27 | 733.29 | 9.21 | 6.73 | 13.79 |
| R30 | 410.62 | 2378.92 | 798.63 | 7.41 | 7.28 | 14.69 |
| R31 | 437.43 | 2618.1 | 878.93 | 36.75 | 7.86 | 16.16 |
| R32 | 481.36 | 2885.14 | 968.58 | 7.45 | 8.45 | 18.25 |
| R33 | 543.69 | 3151.48 | 1057.99 | 1.34 | 8.98 | 20.78 |
| R34 | 618.98 | 3380.74 | 1134.96 | 0.2 | 9.41 | 23.29 |
| R35 | 693.81 | 3536.51 | 1187.25 | 3.08 | 9.69 | 25.17 |
| R36 | 749.76 | 4497.42 | 592.02 | 24.33 | 3.27 | 17.08 |
| R37 | 2640.6 | 4796.57 | 597.29 | 21.15 | 3.61 | 19.64 |
| R38 | 2854.56 | 5095.49 | 616.69 | 113.35 | 4.04 | 23.03 |
| R39 | 3084.54 | 5376.41 | 670.68 | 24.68 | 4.53 | 26.36 |
| R40 | 3314.21 | 5620.52 | 782.51 | 4.71 | 5.04 | 28.1 |
| R41 | 3523.72 | 5810 | 949.03 | 0.73 | 5.48 | 27.86 |
| R42 | 3692.38 | 5930.16 | 1112.05 | 11.72 | 5.78 | 26.88 |
| R43 | 3801.98 | 2239.83 | 4346.54 | 0.03 | 3.91 | 18.41 |
| R44 | 1000 | 2195.48 | 4635.65 | 0.03 | 4.16 | 19.64 |
| R45 | 1000 | 2067.63 | 4924.54 | 0.03 | 4.58 | 20.86 |
| R46 | 1000 | 1870.87 | 5196.03 | 0.04 | 5.18 | 22.01 |
| R47 | 1000 | 1626.45 | 5431.95 | 0.04 | 5.9 | 23.01 |
| R48 | 1000 | 1358.53 | 5615.07 | 0.05 | 6.61 | 23.79 |
| R49 | 1000 | 1090.24 | 5731.21 | 0.06 | 7.14 | 24.28 |
| R50 | 1000 | 192.96 | 2218.23 | 0 | 45.43 | 36.08 |
| R51 | 45 | 194.68 | 1957.58 | 0 | 47.02 | 46.51 |
| R52 | 45 | 201 | 1345.43 | 0.01 | 47.2 | 76.36 |
| R53 | 45 | 218.6 | 720.15 | 0.16 | 44.35 | 158.83 |
| R54 | 45 | 255.05 | 300.21 | 6.09 | 37.62 | 416.39 |
| R55 | 45 | 309.32 | 97.46 | 664.64 | 29.04 | 1370.51 |
| R56 | 45 | 362.46 | 24.64 | 1000 | 21.49 | 5647 |
| R57 | 45 | 0.92 | 49.89 | 2.71 | 0.21 | 0.01 |
| R58 | 930 | 30.32 | 23.13 | 4.36 | 1.01 | 14.72 |
| R59 | 564.07 | 369.39 | 805.06 | 0.06 | 0.2 | 0.16 |
| R60 | 125.86 | 1655.47 | 1804.01 | 0.06 | 0.22 | 0.17 |

*continues on next page*

| | | | | | | |
|---|---|---|---|---|---|---|
| R61 | 10.33 | 2729.41 | 2974.31 | 2.46 | 0.21 | 0.02 |
| R62 | 0.31 | 1655.47 | 2146.77 | 1.56 | 0.27 | 0.18 |
| R63 | 0.1 | 1569.84 | 727.85 | 0.51 | 0.26 | 0.11 |
| R64 | 0.1 | 105.06 | 13.64 | 0.73 | 0.37 | 18.97 |
| R65 | 30 | 319.25 | 346.92 | 2.34 | 4.25 | 2.41 |
| R66 | 34 | 103.65 | 112.63 | 7.85 | 0.35 | 0.2 |
| R67 | 37 | 1953.33 | 2653.8 | 0.91 | 21.75 | 56.66 |
| R68 | 194 | 1926.38 | 2890.29 | 0.98 | 20.57 | 60.42 |
| R69 | 216 | 1847.77 | 3180.88 | 1.06 | 17.41 | 64.19 |
| R70 | 74 | 1723.8 | 3505.33 | 1.14 | 13.19 | 67.73 |
| R71 | 374 | 1564.1 | 3828.93 | 1.21 | 8.94 | 70.8 |
| R72 | 26 | 1380.31 | 4107.46 | 1.27 | 5.42 | 73.19 |

Table 6: Estimated relative concentrations of metabolic enzymes in different Arabidopsis ecotypes: *Aa*—0, *Pi*—0, *Wl*—0. $E_t$ stands for the total enzyme concentration. The $E_t$ values in *Aa*—0 ecotype are arbitrarily set to 1.0.

| Parameter | *Aa*—0 | *Pi*—0 | *Wl*—0 |
|---|---|---|---|
| $E_{t,AATR}$ | 1.0 | 4.41 | 1.32 |
| $E_{t,BAT5}$ | 1.0 | 3.22 | 1.07 |
| $E_{t,BCAT3}$ | 1.0 | 1.75 | 0.42 |
| $E_{t,BCAT4}$ | 1.0 | 3.0 | 3.80 |
| $E_{t,CYP79F1}$ | 1.0 | 0.015 | 1.02 |
| $E_{t,CYP79F2}$ | 1.0 | 0.139 | 0.01 |
| $E_{t,MAM1}$ | 1.0 | 0.05 | 0.01 |
| $E_{t,MAM3}$ | 1.0 | 0.73 | 0.84 |

# D

# EXPERIMENTAL DATA USED FOR MODELLING

## D.1 EXPERIMENTAL DATA ON *col* WT

| Profile | 3C | 4C | 5C | 6C | 7C | 8C | Source |
|---|---|---|---|---|---|---|---|
| 1 | 3.8 | 25.4 | 0.6 | 0.2 | 0.9 | 4.5 | [58] |
| 2 | 2.53 | 18.16 | 0.49 | 0.04 | 0.28 | 1.47 | [92] |
| 3 | 2.7 | 16.3 | 0.5 | 0.3 | 0.4 | 2.2 | [97] |
| 4 | 2.87 | 26.03 | 1.2 | 0.15 | 0.56 | 3.11 | [55] |

Table 7: Glucosinolates concentration in the leaves of *Arabidopsis thaliana Col* wild-type ecotype. The shown profiles are the experimentally-measured GSLs (3C . . . 8C) concentration in the leaves of 5-6 weeks old *Arabidopsis thaliana Col* wild-type plants. Quantities are given in $\mu$ mole $g^{-1}$ dry weight. 3C, three carbon GSLs; 4C, four carbon GSLs; 5C, five carbon GSLs; 6C, six carbon GSLs; 7C, seven carbon GSLs; and 8C, eight carbon GSLs.

## D.2 EXPERIMENTAL DATA ON *col* MUTANTS

| Mutant | 3C | 4C | 5C | 6C | 7C | 8C | Source |
|---|---|---|---|---|---|---|---|
| *mam1* | 16.03 | 0.4 | 0.0 | 0.0 | 0.6 | 7.0 | [58] |
| *mam3* | 2.5 | 18.2 | 0.6 | 0.2 | n.d | n.d. | [97] |
| *bcat3* | 3.8 | 25.4 | 0.6 | 0.2 | 0.9 | 4.5 | [55] |
| *bcat4* | 0.7 | 6.73 | 0.0 | 0.0 | 0.26 | 0.64 | [92] |
| *cyp79f1* | 2.53 | 18.16 | 0.49 | 0.04 | 0.28 | 1.47 | [10] |
| *cyp79f2* | 2.7 | 16.3 | 0.5 | 0.3 | 0.4 | 2.2 | [10] |

Table 8: Glucosinolate content in leaves of *Arabidopsis thaliana* mutant backgrounds. Quantities are given in $\mu$ mole $g^{-1}$ dry weight. n.d. stands for not detectable.

| Ecotype | 3C | 4C | 7C | 8C | MAM type |
|---|---|---|---|---|---|
| Ag-0* | 1.53 | 14.34 | 0.31 | 1.82 | MAM1 |
| Bla-10* | 1.03 | 9.91 | 0.27 | 1.53 | MAM1 |
| Bs1* | 1.24 | 11.37 | 0.34 | 1.43 | MAM1 |
| Cal* | 5.06 | 13.32 | 0.57 | 2.29 | MAM1 |
| Cnt* | 0.28 | 22.46 | 0.51 | 3.56 | MAM1 |
| Ema-1* | 0.21 | 16.17 | 0.45 | 2.26 | MAM1 |
| Pog-1* | 0.27 | 20.42 | 0.43 | 2.23 | MAM1 |
| Tac | 0.40 | 16.61 | 0.31 | 1.67 | |
| Kas* | 5.31 | 11.66 | 0.10 | 0.78 | MAM1 |
| Sorbo* | 4.55 | 20.96 | 0.24 | 1.50 | MAM1 |
| Cvi | 10.03 | 24.40 | 0.38 | 2.62 | |
| Di-1* | 1.82 | 4.13 | 0.29 | 1.44 | MAM1 |
| Aa-0* | 0.07 | 2.29 | 0.11 | 1.29 | MAM1 |
| Col* | 0.40 | 5.93 | 0.19 | 1.18 | MAM1 |
| Ma-0 | 0.00 | 2.54 | 0.10 | 0.93 | |
| Mt-0 | 0.00 | 4.28 | 0.15 | 1.28 | |
| Can | 22.70 | 0.67 | 0.33 | 2.77 | |
| Kondara* | 26.14 | 0.47 | 0.14 | 2.42 | MAM2 |
| Ei-2* | 16.79 | 0.23 | 0.20 | 2.62 | MAM1 |
| Hodja | 13.10 | 0.18 | 0.03 | 0.91 | |
| Kil-0* | 14.63 | 0.19 | 0.22 | 2.40 | MAM1 |
| Mrk-0 | 20.42 | 0.29 | 0.18 | 2.68 | |
| Rsch-0 | 10.02 | 0.22 | 0.10 | 1.58 | |
| Su-0* | 14.29 | 0.13 | 0.13 | 2.15 | MAM2 |
| Wl-0* | 14.45 | 0.16 | 0.17 | 1.40 | MAM2 |
| Bl-1* | 7.60 | 0.08 | 0.13 | 2.35 | MAM1 |
| Ka-0 | 14.36 | 0.00 | 0.12 | 1.99 | |
| Ler | 9.24 | 0.02 | 0.10 | 1.65 | |
| Lip-0* | 5.86 | 0.00 | 0.10 | 1.31 | MAM1 |
| No-0 | 7.65 | 0.11 | 0.07 | 1.40 | |
| Pet | 9.38 | 0.09 | 0.12 | 2.19 | |
| Pi-0* | 2.07 | 0.08 | 0.04 | 0.64 | MAM1 |
| Sei-0* | 4.77 | 0.07 | 0.10 | 1.18 | MAM2 |
| Tsu-1 | 6.02 | 0.11 | 0.14 | 1.53 | |
| Yo-0* | 6.01 | 0.21 | 0.11 | 1.16 | MAM2 |

Table 10: Glucosinolate contents in the leaves of *Arabidopsis* ecotypes. Quantities are given in μmol g dry $wt^{-1}$. Ecotypes are marked with *, where information on both amino acid sequences of metabolic genes and GSL concentrations are known.

# E

# SHANNON ENTROPY OF GLUCOSINOLATE GENES

Table 11: Average shannon entropy across the length of glucosinolate genes of *Arabidospsis thaliana* ecotypes taken from [1] and [52]. ASE1, average shannon entropy of 343 ecotypes from [1]; ASE2, average shannon entropy of 22 ecotypes from [52].

| Glucosinolate Gene | Average shannon entropy | |
|---|---|---|
| | ASE1 | ASE2 |
| SOT17 | 0.000486 | 0.000363 |
| BCAT4 | 0.000704 | 0.000355 |
| CYP79B2 | 0.000619 | 0.000307 |
| GSTF11 | 0.002284 | 0.002487 |
| CYP79A2 | 0.002308 | 0.001961 |
| BCAT3 | 0.001281 | 0.000631 |
| MAM3 | 0.002427 | 0.002296 |
| FMO-GSOX3 | 0.005172 | 0.004464 |
| GSTF10 | 0.0 | 0.0 |
| UGT74C1 | 0.001672 | 0.000890 |
| GSTU20 | 0.002062 | 0.001937 |
| APK1 | 0.000872 | 0.001545 |
| IPMI SSU2 | 0.001638 | 0.001441 |
| CYP79F1 | 0.002459 | 0.001729 |
| GS-OH | 0.002389 | 0.002724 |
| AOP3 | 0.003192 | 0.005107 |
| GSH1 | 0.000817 | 0.000871 |
| GGP1 | 0.000466 | 0.000251 |
| IPMI SSU3 | 0.001810 | 0.001773 |
| UGT74B1 | 0.006034 | 0.005173 |
| BAT5 | 0.000169 | 0.000460 |

*continues on next page*

| | | |
|---|---|---|
| CYP83B1 | 0.0 | 0.000126 |
| CYP79F2 | 0.001400 | 0.000980 |
| AAO4 | 0.002284 | 0.001910 |
| SOT18 | 0.001525 | 0.001611 |
| SOT16 | 0.000779 | 0.000744 |
| CYP83A1 | 0.000263 | 0.000250 |
| FMO-GSOX2 | 0.009534 | 0.007247 |
| CHY1 | 0.001987 | 0.000191 |
| IPMDH1 | 0.001090 | 0.001212 |
| IPMI LSU1 | 0.000397 | 0.000247 |
| BZO1 | 0.001638 | 0.001122 |
| FMO-GSOX4 | 0.001699 | 0.001059 |
| FMO-GSOX5 | 0.000857 | 0.000429 |
| CYP81F2 | 0.002731 | 0.002880 |
| GSTF9 | 0.000116 | 0.0 |
| IPMDH3 | 0.002906 | 0.001930 |
| FMO-GSOX1 | 0.001665 | 0.001095 |
| MAM1 | 0.009653 | 0.007920 |
| SUR1 | 0.005192 | 0.004651 |
| CYP79B3 | 0.002359 | 0.002245 |

# F

# SUPPLEMENTARY MATERIALS ABOUT THE PHYLOGENETIC ANALYSIS

Table 12: Evolutionary distances based on sequence similarity to MAM1 and MAM2 genes from 343 ecotypes [1]. The values are substitutions per site.

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 9602 | 0.08784039 | 0.02538253 | MAM2 |
| 9547 | 0.0878203 | 0.02536244 | MAM2 |
| 9789 | 0.08430283 | 0.02184497 | MAM2 |
| 9589 | 0.08430284 | 0.02184498 | MAM2 |
| 9588 | 0.08430285 | 0.02184499 | MAM2 |
| 9584 | 0.08430286 | 0.021845 | MAM2 |
| 9577 | 0.08430287 | 0.02184501 | MAM2 |
| 9567 | 0.08430288 | 0.02184502 | MAM2 |
| 9564 | 0.08430289 | 0.02184503 | MAM2 |
| 9558 | 0.0843029 | 0.02184504 | MAM2 |
| 9557 | 0.08430291 | 0.02184505 | MAM2 |
| 9546 | 0.08430292 | 0.02184506 | MAM2 |
| 9592 | 0.09142665 | 0.02896879 | MAM2 |
| 9520 | 0.08430298 | 0.02184512 | MAM2 |
| 9525 | 0.0843031 | 0.02184524 | MAM2 |
| 9531 | 0.08430335 | 0.02184549 | MAM2 |
| 2278 | 0.08430207 | 0.02184421 | MAM2 |
| 1853 | 0.08430193 | 0.02184407 | MAM2 |
| 932 | 0.0914688 | 0.02901094 | MAM2 |
| 4931 | 0.08430175 | 0.02184389 | MAM2 |
| 5768 | 0.08430146 | 0.0218436 | MAM2 |
| 9534 | 0.08782646 | 0.0253692 | MAM2 |
| 9594 | 0.10211517 | 0.03965791 | MAM2 |
| 6744 | 0.09135594 | 0.02889868 | MAM2 |
| 7947 | 0.08430097 | 0.02184373 | MAM2 |
| 8132 | 0.08430085 | 0.02184385 | MAM2 |
| 870 | 0.08430056 | 0.02184416 | MAM2 |
| 9572 | 0.07716042 | 0.02898428 | MAM2 |
| 9932 | 0.07716041 | 0.02898429 | MAM2 |
| 9723 | 0.05495774 | 0.05118724 | MAM2 |
| 9576 | 0.05486184 | 0.0654996 | MAM1 |
| 9556 | 0.05486201 | 0.06549977 | MAM1 |
| 9580 | 0.05486202 | 0.06549978 | MAM1 |
| 9581 | 0.05486203 | 0.06549979 | MAM1 |
| 9637 | 0.04033725 | 0.06580797 | MAM1 |
| 9553 | 0.02920676 | 0.0769385 | MAM1 |
| 2171 | 0.02555834 | 0.08058686 | MAM1 |
| 265 | 0.02555785 | 0.08058685 | MAM1 |
| 403 | 0.02555773 | 0.08058697 | MAM1 |
| 424 | 0.02555748 | 0.08058722 | MAM1 |
| 5104 | 0.02555727 | 0.08058743 | MAM1 |
| 5165 | 0.02555701 | 0.08058769 | MAM1 |

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 6390 | 0.02555688 | 0.08058782 | MAM1 |
| 6434 | 0.02555681 | 0.08058789 | MAM1 |
| 7296 | 0.02555678 | 0.08058792 | MAM1 |
| 9924 | 0.02922231 | 0.08425409 | MAM1 |
| 9783 | 0.02922232 | 0.0842541 | MAM1 |
| 9782 | 0.02922233 | 0.08425411 | MAM1 |
| 9601 | 0.02922234 | 0.08425412 | MAM1 |
| 9591 | 0.02922235 | 0.08425413 | MAM1 |
| 9587 | 0.02922236 | 0.08425414 | MAM1 |
| 9586 | 0.02922237 | 0.08425415 | MAM1 |
| 9532 | 0.02922238 | 0.08425416 | MAM1 |
| 9529 | 0.02922239 | 0.08425417 | MAM1 |
| 5151 | 0.0292224 | 0.08425418 | MAM1 |
| 5784 | 0.0292224 | 0.08425418 | MAM1 |
| 9718 | 0.03653978 | 0.09157216 | MAM1 |
| 9757 | 0.03653978 | 0.09157216 | MAM1 |
| 9524 | 0.03292223 | 0.08795461 | MAM1 |
| 9540 | 0.03292237 | 0.08795475 | MAM1 |
| 9568 | 0.03292251 | 0.08795489 | MAM1 |
| 9748 | 0.03658714 | 0.09161952 | MAM1 |
| 9719 | 0.03658715 | 0.09161953 | MAM1 |
| 9539 | 0.03658716 | 0.09161954 | MAM1 |
| 9716 | 0.04379787 | 0.09883025 | MAM1 |
| 9714 | 0.03658717 | 0.09161955 | MAM1 |
| 9639 | 0.03288362 | 0.087916 | MAM1 |
| 9125 | 0.029198 | 0.08423038 | MAM1 |
| 9662 | 0.02919817 | 0.08423055 | MAM1 |
| 9638 | 0.02919818 | 0.08423056 | MAM1 |
| 9636 | 0.02919819 | 0.08423057 | MAM1 |
| 9624 | 0.0291982 | 0.08423058 | MAM1 |
| 9563 | 0.02919821 | 0.08423059 | MAM1 |
| 1890 | 0.02919822 | 0.0842306 | MAM1 |
| 9527 | 0.029202 | 0.08423494 | MAM1 |
| 9133 | 0.02920251 | 0.08423545 | MAM1 |
| 9543 | 0.02920238 | 0.08423532 | MAM1 |
| 9745 | 0.02920267 | 0.08423561 | MAM1 |
| 9617 | 0.02920268 | 0.08423562 | MAM1 |
| 9609 | 0.02920269 | 0.08423563 | MAM1 |
| 9134 | 0.02920284 | 0.08423578 | MAM1 |
| 9627 | 0.03284003 | 0.08787297 | MAM1 |
| 9919 | 0.03645794 | 0.09149088 | MAM1 |
| 9635 | 0.01824216 | 0.08790278 | MAM1 |
| 7158 | 0.02187712 | 0.091538 | MAM1 |
| 9608 | 0.01822533 | 0.09523297 | MAM1 |
| 9665 | 0.01456878 | 0.09157642 | MAM1 |
| 9535 | 0.01091336 | 0.10250538 | MAM1 |
| 9593 | 0.00727631 | 0.09886891 | MAM1 |
| 1925 | 0.00365802 | 0.10248668 | MAM1 |
| 1954 | 0.00365801 | 0.10248669 | MAM1 |
| 9128 | 0.003658 | 0.1024867 | MAM1 |
| 9130 | 0.00365799 | 0.10248671 | MAM1 |
| 9131 | 0.00365798 | 0.10248672 | MAM1 |
| 9744 | 0.00365797 | 0.10248673 | MAM1 |
| 9933 | 0.00365796 | 0.10248674 | MAM1 |
| MAM1 | 0 | 0.10614468 | MAM1 |
| 108 | 3.00E-08 | 0.10614469 | MAM1 |
| 139 | 4.00E-08 | 0.1061447 | MAM1 |
| 2276 | 5.00E-08 | 0.10614471 | MAM1 |
| 351 | 6.00E-08 | 0.10614472 | MAM1 |
| 4779 | 7.00E-08 | 0.10614473 | MAM1 |
| 4807 | 8.00E-08 | 0.10614474 | MAM1 |

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 4826 | 9.00E-08 | 0.10614475 | MAM1 |
| 4958 | 1.00E-07 | 0.10614476 | MAM1 |
| 5253 | 1.10E-07 | 0.10614477 | MAM1 |
| 5276 | 1.20E-07 | 0.10614478 | MAM1 |
| 5779 | 1.30E-07 | 0.10614479 | MAM1 |
| 5818 | 1.40E-07 | 0.1061448 | MAM1 |
| 630 | 1.50E-07 | 0.10614481 | MAM1 |
| 7015 | 1.60E-07 | 0.10614482 | MAM1 |
| 8266 | 1.70E-07 | 0.10614483 | MAM1 |
| 8411 | 1.80E-07 | 0.10614484 | MAM1 |
| 8472 | 1.90E-07 | 0.10614485 | MAM1 |
| 88 | 2.00E-07 | 0.10614486 | MAM1 |
| 9100 | 2.10E-07 | 0.10614487 | MAM1 |
| 9102 | 2.20E-07 | 0.10614488 | MAM1 |
| 9312 | 2.30E-07 | 0.10614489 | MAM1 |
| 9314 | 2.40E-07 | 0.1061449 | MAM1 |
| 9507 | 2.50E-07 | 0.10614491 | MAM1 |
| 9510 | 2.60E-07 | 0.10614492 | MAM1 |
| 9511 | 2.70E-07 | 0.10614493 | MAM1 |
| 9512 | 2.80E-07 | 0.10614494 | MAM1 |
| 9514 | 2.90E-07 | 0.10614495 | MAM1 |
| 9515 | 3.00E-07 | 0.10614496 | MAM1 |
| 9521 | 3.10E-07 | 0.10614497 | MAM1 |
| 9522 | 0.003585 | 0.10972966 | MAM1 |
| 9526 | 3.70E-07 | 0.10614503 | MAM1 |
| 9537 | 3.90E-07 | 0.10614505 | MAM1 |
| 9538 | 4.00E-07 | 0.10614506 | MAM1 |
| 9544 | 4.10E-07 | 0.10614507 | MAM1 |
| 9559 | 4.20E-07 | 0.10614508 | MAM1 |
| 9560 | 4.30E-07 | 0.10614509 | MAM1 |
| 9571 | 4.40E-07 | 0.1061451 | MAM1 |
| 9590 | 4.50E-07 | 0.10614511 | MAM1 |
| 9606 | 4.60E-07 | 0.10614512 | MAM1 |
| 9646 | 4.70E-07 | 0.10614513 | MAM1 |
| 9649 | 4.80E-07 | 0.10614514 | MAM1 |
| 9650 | 4.90E-07 | 0.10614515 | MAM1 |
| 9651 | 5.00E-07 | 0.10614516 | MAM1 |
| 9652 | 5.10E-07 | 0.10614517 | MAM1 |
| 9653 | 5.20E-07 | 0.10614518 | MAM1 |
| 9655 | 5.30E-07 | 0.10614519 | MAM1 |
| 9657 | 5.40E-07 | 0.1061452 | MAM1 |
| 9658 | 5.50E-07 | 0.10614521 | MAM1 |
| 9661 | 5.60E-07 | 0.10614522 | MAM1 |
| 9698 | 5.70E-07 | 0.10614523 | MAM1 |
| 9702 | 5.80E-07 | 0.10614524 | MAM1 |
| 9725 | 5.90E-07 | 0.10614525 | MAM1 |
| 9726 | 6.00E-07 | 0.10614526 | MAM1 |
| 9746 | 6.10E-07 | 0.10614527 | MAM1 |
| 9809 | 6.20E-07 | 0.10614528 | MAM1 |
| 9908 | 6.30E-07 | 0.10614529 | MAM1 |
| 9910 | 6.40E-07 | 0.1061453 | MAM1 |
| 9911 | 6.50E-07 | 0.10614531 | MAM1 |
| 9917 | 6.60E-07 | 0.10614532 | MAM1 |
| 9920 | 6.70E-07 | 0.10614533 | MAM1 |
| 9922 | 6.80E-07 | 0.10614534 | MAM1 |
| 9923 | 6.90E-07 | 0.10614535 | MAM1 |
| 9925 | 7.00E-07 | 0.10614536 | MAM1 |
| 9926 | 7.10E-07 | 0.10614537 | MAM1 |
| 9927 | 7.20E-07 | 0.10614538 | MAM1 |
| 9929 | 7.30E-07 | 0.10614539 | MAM1 |
| 9574 | 0.00358633 | 0.10973099 | MAM1 |

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 9555 | 0.00358634 | 0.109731 | MAM1 |
| 9554 | 0.00358635 | 0.10973101 | MAM1 |
| 9545 | 0.00358636 | 0.10973102 | MAM1 |
| 9541 | 0.00358637 | 0.10973103 | MAM1 |
| 9533 | 0.00358638 | 0.10973104 | MAM1 |
| 9530 | 0.00358639 | 0.10973105 | MAM1 |
| 9518 | 0.0035864 | 0.10973106 | MAM1 |
| 7418 | 0.00358641 | 0.10973107 | MAM1 |
| 9569 | 0.01821266 | 0.12435732 | MAM1 |
| 9600 | 0.01079627 | 0.11694093 | MAM1 |
| 5772 | 0.01084111 | 0.11698577 | MAM1 |
| 9928 | 0.01084125 | 0.11698591 | MAM1 |
| 9918 | 0.01084126 | 0.11698592 | MAM1 |
| 9704 | 0.01084127 | 0.11698593 | MAM1 |
| 9699 | 0.01084128 | 0.11698594 | MAM1 |
| 9697 | 0.01084129 | 0.11698595 | MAM1 |
| 9597 | 0.0108413 | 0.11698596 | MAM1 |
| 9585 | 0.01084131 | 0.11698597 | MAM1 |
| 9579 | 0.01084132 | 0.11698598 | MAM1 |
| 9575 | 0.01084133 | 0.11698599 | MAM1 |
| 9573 | 0.01084134 | 0.116986 | MAM1 |
| 9570 | 0.01084135 | 0.11698601 | MAM1 |
| 9566 | 0.01084136 | 0.11698602 | MAM1 |
| 9565 | 0.01084137 | 0.11698603 | MAM1 |
| 9562 | 0.01084138 | 0.11698604 | MAM1 |
| 9523 | 0.01084139 | 0.11698605 | MAM1 |
| 9517 | 0.0108414 | 0.11698606 | MAM1 |
| 9508 | 0.01084141 | 0.11698607 | MAM1 |
| 9506 | 0.01084142 | 0.11698608 | MAM1 |
| 9104 | 0.01084143 | 0.11698609 | MAM1 |
| 350 | 0.01084144 | 0.1169861 | MAM1 |
| 9934 | 9.00E-07 | 0.10614556 | MAM1 |
| 9938 | 1.06E-06 | 0.10614572 | MAM1 |
| 9937 | 0.0035964 | 0.10974106 | MAM1 |
| 9931 | 0.00359641 | 0.10974107 | MAM1 |
| 9913 | 0.00359642 | 0.10974108 | MAM1 |
| 9909 | 0.00359644 | 0.1097411 | MAM1 |
| 5210 | 0.00359653 | 0.10974119 | MAM1 |
| 5023 | 0.00359664 | 0.1097413 | MAM1 |
| 4900 | 0.00359685 | 0.10974151 | MAM1 |
| 4884 | 0.00359702 | 0.10974168 | MAM1 |
| 2016 | 0.00359715 | 0.10974181 | MAM1 |
| 9663 | 0.00717825 | 0.11332291 | MAM1 |
| 9660 | 0.00359648 | 0.10974114 | MAM1 |
| 9659 | 0.00359649 | 0.10974115 | MAM1 |
| 9519 | 0.0035965 | 0.10974116 | MAM1 |
| 9503 | 0.00359651 | 0.10974117 | MAM1 |
| 5726 | 0.00359652 | 0.10974118 | MAM1 |
| 5644 | 0.00359653 | 0.10974119 | MAM1 |
| 5577 | 0.00359654 | 0.1097412 | MAM1 |
| 5236 | 0.00359655 | 0.10974121 | MAM1 |
| 5353 | 0.00359655 | 0.10974121 | MAM1 |
| 9914 | 0.02555677 | 0.08058795 | MAM1 |
| 9907 | 0.02555678 | 0.08058796 | MAM1 |
| 9816 | 0.02555679 | 0.08058797 | MAM1 |
| 9815 | 0.0255568 | 0.08058798 | MAM1 |
| 9810 | 0.02555681 | 0.08058799 | MAM1 |
| 9806 | 0.02555682 | 0.080588 | MAM1 |
| 9799 | 0.02555683 | 0.08058801 | MAM1 |
| 9797 | 0.02555684 | 0.08058802 | MAM1 |
| 9794 | 0.02555685 | 0.08058803 | MAM1 |

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 9792 | 0.02555686 | 0.08058804 | MAM1 |
| 9774 | 0.02555687 | 0.08058805 | MAM1 |
| 9771 | 0.02555688 | 0.08058806 | MAM1 |
| 9769 | 0.02555689 | 0.08058807 | MAM1 |
| 9737 | 0.0255569 | 0.08058808 | MAM1 |
| 9727 | 0.02555691 | 0.08058809 | MAM1 |
| 9669 | 0.02555692 | 0.0805881 | MAM1 |
| 9668 | 0.02555693 | 0.08058811 | MAM1 |
| 9667 | 0.02555694 | 0.08058812 | MAM1 |
| 9654 | 0.02555695 | 0.08058813 | MAM1 |
| 9643 | 0.02555696 | 0.08058814 | MAM1 |
| 9642 | 0.02555697 | 0.08058815 | MAM1 |
| 9640 | 0.02555698 | 0.08058816 | MAM1 |
| 9633 | 0.02555699 | 0.08058817 | MAM1 |
| 9632 | 0.025557 | 0.08058818 | MAM1 |
| 9630 | 0.02555701 | 0.08058819 | MAM1 |
| 9629 | 0.02555702 | 0.0805882 | MAM1 |
| 9628 | 0.02555703 | 0.08058821 | MAM1 |
| 9626 | 0.02555704 | 0.08058822 | MAM1 |
| 9625 | 0.02555705 | 0.08058823 | MAM1 |
| 9620 | 0.02555706 | 0.08058824 | MAM1 |
| 9619 | 0.02555707 | 0.08058825 | MAM1 |
| 9615 | 0.02555708 | 0.08058826 | MAM1 |
| 9612 | 0.02555709 | 0.08058827 | MAM1 |
| 9607 | 0.0255571 | 0.08058828 | MAM1 |
| 9599 | 0.02555711 | 0.08058829 | MAM1 |
| 9596 | 0.02555712 | 0.0805883 | MAM1 |
| 9595 | 0.02555713 | 0.08058831 | MAM1 |
| 9578 | 0.02555714 | 0.08058832 | MAM1 |
| 9561 | 0.02555715 | 0.08058833 | MAM1 |
| 9552 | 0.02555716 | 0.08058834 | MAM1 |
| 8428 | 0.02555717 | 0.08058835 | MAM1 |
| 9550 | 0.02555717 | 0.08058835 | MAM1 |
| 9713 | 0.09142973 | 0.02185727 | MAM2 |
| 9721 | 0.08785888 | 0.01828642 | MAM2 |
| MAM2 | 0.10614468 | 0 | MAM2 |
| 8077 | 0.10211924 | 0.02546664 | MAM2 |
| 9739 | 0.11648327 | 0.03983067 | MAM2 |
| 9813 | 0.10918339 | 0.03253079 | MAM2 |
| 9712 | 0.11270204 | 0.03604944 | MAM2 |
| 9645 | 0.10918381 | 0.03253121 | MAM2 |
| 9733 | 0.11623687 | 0.03958427 | MAM2 |
| 9743 | 0.11981574 | 0.04316314 | MAM2 |
| 9741 | 0.11623849 | 0.03958589 | MAM2 |
| 428 | 0.11269986 | 0.03604726 | MAM2 |
| 5907 | 0.11269987 | 0.03604727 | MAM2 |
| 410 | 0.11974177 | 0.04308917 | MAM2 |
| 5890 | 0.11621853 | 0.03956593 | MAM2 |
| 9930 | 0.11270034 | 0.03604774 | MAM2 |
| 9916 | 0.11270035 | 0.03604775 | MAM2 |
| 9915 | 0.11270036 | 0.03604776 | MAM2 |
| 9811 | 0.11270037 | 0.03604777 | MAM2 |
| 9807 | 0.11270038 | 0.03604778 | MAM2 |
| 9805 | 0.11270039 | 0.03604779 | MAM2 |
| 9804 | 0.1127004 | 0.0360478 | MAM2 |
| 9803 | 0.11270041 | 0.03604781 | MAM2 |
| 9801 | 0.11270042 | 0.03604782 | MAM2 |
| 9800 | 0.11270043 | 0.03604783 | MAM2 |
| 9798 | 0.11270044 | 0.03604784 | MAM2 |
| 9796 | 0.11270045 | 0.03604785 | MAM2 |
| 9791 | 0.11270046 | 0.03604786 | MAM2 |

| Ecotype ID | dMAM1 | dMAM2 | MAM type |
|---|---|---|---|
| 9790 | 0.11270047 | 0.03604787 | MAM2 |
| 9788 | 0.11270048 | 0.03604788 | MAM2 |
| 9786 | 0.11270049 | 0.03604789 | MAM2 |
| 9785 | 0.1127005 | 0.0360479 | MAM2 |
| 9784 | 0.11270051 | 0.03604791 | MAM2 |
| 9781 | 0.11270052 | 0.03604792 | MAM2 |
| 9780 | 0.11270053 | 0.03604793 | MAM2 |
| 9779 | 0.11270054 | 0.03604794 | MAM2 |
| 9778 | 0.11270055 | 0.03604795 | MAM2 |
| 9777 | 0.11270056 | 0.03604796 | MAM2 |
| 9776 | 0.11270057 | 0.03604797 | MAM2 |
| 9775 | 0.11270058 | 0.03604798 | MAM2 |
| 9772 | 0.11270059 | 0.03604799 | MAM2 |
| 9770 | 0.1127006 | 0.036048 | MAM2 |
| 9768 | 0.11270061 | 0.03604801 | MAM2 |
| 9749 | 0.11270062 | 0.03604802 | MAM2 |
| 9730 | 0.11270063 | 0.03604803 | MAM2 |
| 9728 | 0.11270064 | 0.03604804 | MAM2 |
| 9644 | 0.11270065 | 0.03604805 | MAM2 |
| 9622 | 0.11270066 | 0.03604806 | MAM2 |
| 915 | 0.11270067 | 0.03604807 | MAM2 |
| 801 | 0.11270068 | 0.03604808 | MAM2 |
| 6445 | 0.11270069 | 0.03604809 | MAM2 |
| 6424 | 0.1127007 | 0.0360481 | MAM2 |
| 6396 | 0.11270071 | 0.03604811 | MAM2 |
| 6296 | 0.11270072 | 0.03604812 | MAM2 |
| 5993 | 0.11270073 | 0.03604813 | MAM2 |
| 5984 | 0.11270076 | 0.03604816 | MAM2 |
| 5950 | 0.11270081 | 0.03604821 | MAM2 |
| 5874 | 0.11624406 | 0.03959146 | MAM2 |
| 9736 | 0.11620074 | 0.03954814 | MAM2 |
| 5811 | 0.11622692 | 0.03957432 | MAM2 |
| 7477 | 0.11626203 | 0.03960943 | MAM2 |
| 1829 | 0.11270197 | 0.03604937 | MAM2 |
| 9814 | 0.11620437 | 0.03955177 | MAM2 |
| 1872 | 0.11270232 | 0.03604972 | MAM2 |
| 5893 | 0.11269929 | 0.03604669 | MAM2 |
| 5800 | 0.11269942 | 0.03604682 | MAM2 |
| 9793 | 0.11621172 | 0.03955912 | MAM2 |
| 7346 | 0.12330107 | 0.04664847 | MAM2 |
| 9808 | 0.1233009 | 0.0466483 | MAM2 |
| 7231 | 0.12328998 | 0.04663738 | MAM2 |
| 9598 | 0.1057551 | 0.0291025 | MAM2 |
| 9634 | 0.10569533 | 0.02904273 | MAM2 |
| 9551 | 0.09497428 | 0.01832168 | MAM2 |
| 9706 | 0.0985351 | 0.0218825 | MAM2 |
| 9696 | 0.09497466 | 0.01832206 | MAM2 |
| 9695 | 0.09854056 | 0.02188796 | MAM2 |
| 9583 | 0.0949744 | 0.0183218 | MAM2 |
| 9542 | 0.10226349 | 0.02561089 | MAM2 |
| 9582 | 0.09497471 | 0.01832211 | MAM2 |
| 9666 | 0.10216021 | 0.02550761 | MAM2 |
| 9681 | 0.109342 | 0.0326894 | MAM2 |
| 9528 | 0.09856749 | 0.02191489 | MAM2 |
| 9610 | 0.11338273 | 0.03673013 | MAM2 |
| 9611 | 0.11338286 | 0.03673026 | MAM2 |
| 9549 | 0.11338287 | 0.03673027 | MAM2 |
| 159 | 0.09141115 | 0.02183843 | MAM2 |
| 9671 | 0.09494334 | 0.02537062 | MAM2 |
| 9613 | 0.08785886 | 0.01828614 | MAM2 |
| 9536 | 0.08785887 | 0.01828615 | MAM2 |

# BIBLIOGRAPHY


[1] *1001genomes*. URL: http://1001genomes.org/projects/MPICWang2013/index.html (visited on 01/22/2018).

[2] Niels Agerbirk, Carl Erik Olsen, Christine Heimes, Stina Christensen, Søren Bak, and Thure P. Hauser. "Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris." In: *Phytochemistry* 115 (2015), pp. 130–142. ISSN: 00319422. DOI: 10.1016/j.phytochem.2014.09.003. URL: http://www.ncbi.nlm.nih.gov/pubmed/25277803http://linkinghub.elsevier.com/retrieve/pii/S0031942214003665.

[3] Søren Bak, Frans E Tax, Kenneth A Feldmann, David W Galbraith, and René Feyereisen. "CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in Arabidopsis." In: *The Plant Cell* 13.1 (2001), pp. 101–111.

[4] Paweł Bednarek, Mariola Piślewska-Bednarek, Aleš Svatoš, Bernd Schneider, Jan Doubskỳ, Madina Mansurova, Matt Humphry, Chiara Consonni, Ralph Panstruga, Andrea Sanchez-Vallet, et al. "A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense." In: *Science* 323.5910 (2009), pp. 101–106.

[5] Markus Benderoth, Susanne Textor, Aaron J Windsor, Thomas Mitchell-Olds, Jonathan Gershenzon, and Juergen Kroymann. "Positive selection driving diversification in plant secondary metabolism." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.24 (2006), pp. 9118–23. ISSN: 0027-8424. DOI: 10.1073/pnas.0601738103. URL: http://www.ncbi.nlm.nih.gov/pubmed/16754868http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1482576.

[6] G Brader, Eva Tas E, and E T Palva. "Jasmonate-dependent induction of indole glucosinolates in Arabidopsis by culture filtrates of the nonspecific pathogen Erwinia carotovora." In: *Plant physiology* 126.2 (2001), pp. 849–60. ISSN: 0032-0889. URL: http://www.ncbi.nlm.nih.gov/pubmed/11402212http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC111174.

[7] Paul D Brown, Jim G Tokuhisa, Michael Reichelt, and Jonathan Gershenzon. "Variation of glucosinolate accumulation among different organs and developmental stages of

Arabidopsis thaliana." In: *Phytochemistry* 62.3 (2003), pp. 471–481.

[8] S Buskov, B Serra, E Rosa, H Sørensen, and J C Sørensen. "Effects of intact glucosinolates and products produced from glucosinolates in myrosinase-catalyzed hydrolysis on the potato cyst nematode (Globodera rostochiensis Cv. Woll)." In: *Journal of agricultural and food chemistry* 50.4 (2002), pp. 690–5. ISSN: 0021-8561. URL: http://www.ncbi.nlm.nih.gov/pubmed/11829629.

[9] María Elena Cartea and Pablo Velasco. "Glucosinolates in Brassica foods: bioavailability in food and significance for human health." In: *Phytochemistry Reviews* 7.2 (2008), pp. 213–229. ISSN: 1568-7767. DOI: 10.1007/s11101-007-9072-2. URL: http://link.springer.com/10.1007/s11101-007-9072-2.

[10] Sixue Chen, Erich Glawischnig, Kirsten Jørgensen, Peter Naur, Bodil Jørgensen, Carl-Erik Erik Olsen, Carsten H. Hansen, Hasse Rasmussen, John a. Pickett, and Barbara a. Halkier. "CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis." In: *Plant Journal* 33.5 (2003), pp. 923–937. ISSN: 09607412. DOI: 10.1046/j.1365-313X.2003.01679.x. URL: http://doi.wiley.com/10.1046/j.1365-313X.2003.01679.x.

[11] Bo Cheng, Min Lin, Guoyou Huang, Yuhui Li, Baohua Ji, Guy M Genin, Vikram S Deshpande, Tian Jian Lu, and Feng Xu. "Cellular mechanosensing of the biophysical microenvironment: A review of mathematical models of biophysical regulation of cell responses." In: *Physics of life reviews* (2017).

[12] M. D. Chisholm and L. R. Wetter. "BIOSYNTHESIS OF MUSTARD OIL GLUCOSIDES: IV. THE ADMINISTRATION OF METHIONINE-C14 AND RELATED COMPOUNDS TO HORSERADISH." In: *Canadian Journal of Biochemistry* 42.7 (1964), pp. 1033–1040. ISSN: 0008-4018. DOI: 10.1139/o64-114. URL: http://www.nrcresearchpress.com/doi/abs/10.1139/o64-114.

[13] T. C. Chou and P. Talalay. "A simple generalized equation for the analysis of multiple inhibitions of Michaelis-Menten kinetic systems." In: *Journal of Biological Chemistry* 252.18 (1977), pp. 6438–6442. ISSN: 00219258. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-0017706693{\&}partnerID=tZOtx3y1.

[14] UniProt Consortium. "UniProt: the universal protein knowledgebase." In: *Nucleic acids research* 45.D1 (2016), pp. D158–D169.

[15] Melvin E Daxenbichler, Gayland F Spencer, Diana G Carlson, Gertrude B Rose, Anita M Brinker, and Richard G Powell. "Glucosinolate composition of seeds from 297 species of wild plants." In: *Phytochemistry* 30.8 (1991), pp. 2623–2638.

[16] Hidde De Jong. "Modeling and simulation of genetic regulatory systems: a literature review." In: *Journal of computational biology* 9.1 (2002), pp. 67–103.

[17] C Douglas Grubb, Brandon J Zipp, Jutta Ludwig-Müller, Makoto N Masuno, Tadeusz F Molinski, and Steffen Abel. "Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis." In: *The Plant Journal* 40.6 (2004), pp. 893–908.

[18] Patrick P Edger, Hanna M Heidel-Fischer, Michaël Bekaert, Jadranka Rota, Gernot Glöckner, Adrian E Platts, David G Heckel, Joshua P Der, Eric K Wafula, Michelle Tang, et al. "The butterfly plant arms-race escalated by gene and genome duplications." In: *Proceedings of the National Academy of Sciences* 112.27 (2015), pp. 8362–8366.

[19] MG Ettlinger and A Kjaer. "Sulfur compounds in plants." In: *Recent advances in phytochemistry* 1.59 (1968), p. 144.

[20] Martin G. Ettlinger and Allan J. Lundeen. "The Mustard Oil of Limnanthes douglasii Seed, m-Methoxybenzyl Isothiocyanate." In: *Journal of the American Chemical Society* 78.9 (1956), pp. 1952–1954. ISSN: 0002-7863. DOI: 10.1021/ja01590a052. URL: http://pubs.acs.org/doi/abs/10.1021/ja01590a052.

[21] **Karen van Eunen, Sereh MJ Simons, Albert Gerding, Aycha Bleeker, Gijs den Besten, Catharina ML Touw, Sander M Houten, Bert K Groen, Klaas Krab, Dirk-Jan Reijngoud, et al. "Biochemical competition makes fatty-acid β-oxidation vulnerable to substrate overload." In: *PLoS computational biology* 9.8 (2013). e1003186.**

*[21] Substrate overload*

[22] J W Fahey, a T Zalcmann, and P Talalay. "The chemical diversity and distribution of glucosinolates and isothiocyanates amoung plants." In: *Phytochemistry* 56 (2001), pp. 5–51.

[23] David A Fell and Keith Snell. "Control analysis of mammalian serine biosynthesis. Feedback inhibition on the final step." In: *Biochemical Journal* 256.1 (1988), p. 97.

[24] G Roger Fenwick, Robert K Heaney, W John Mullin, and Cecil H VanEtten. "Glucosinolates and their breakdown products in food and food plants." In: *CRC Critical Reviews in Food Science and Nutrition* 18.2 (1983), pp. 123–201.

[25] R. A. Fisher. "On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P." In: *Journal of the Royal Statistical Society* 85.1 (1922), p. 87. ISSN: 09528385. DOI: 10.2307/2340521. URL: http://www.jstor.org/stable/2340521?origin=crossref.

[26] R A Fisher. "GENETICAL THEORY OF NATURAL SELECTION." In: (1930). URL: http://14.139.56.90/bitstream/1/2033620/1/IVRI3205.pdf.

[27] D GARFINKEL and B HESS. "METABOLIC CONTROL MECHANISMS. VII.A DETAILED COMPUTER MODEL OF THE GLYCOLYTIC PATHWAY IN ASCITES CELLS." In: *The Journal of biological chemistry* 239 (1964), pp. 971–83. ISSN: 0021-9258. URL: http://www.ncbi.nlm.nih.gov/pubmed/14165947.

[28] B. Gabrys and W.F. Tjallingii. "The role of sinigrin in host plant recognition by aphids during initial plant penetration." In: *Entomologia Experimentalis et Applicata* 104.1 (2002), pp. 89–93. ISSN: 0013-8703. DOI: 10.1046/j.1570-7458.2002.00994.x. URL: http://doi.wiley.com/10.1046/j.1570-7458.2002.00994.x.

[29] Claire MM Gachon, Mathilde Langlois-Meurinne, Yves Henry, and Patrick Saindrenan. "Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications." In: *Plant molecular biology* 58.2 (2005), pp. 229–245.

[30] Tamara Gigolashvili, Ruslan Yatusevich, Inga Rollwitz, Melanie Humphry, Jonathan Gershenzon, and Ulf-Ingo Flügge. "The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in Arabidopsis thaliana." In: *The Plant cell* 21.6 (2009), pp. 1813–1829. ISSN: 1040-4651. DOI: 10.1105/tpc.109.066399.

[31] Anne Louise Gimsing and John A. Kirkegaard. "Glucosinolates and biofumigation: fate of glucosinolates and their hydrolysis products in soil." In: *Phytochemistry Reviews* 8.1 (2009), pp. 299–310. ISSN: 1568-7767. DOI: 10.1007/s11101-008-9105-5. URL: http://link.springer.com/10.1007/s11101-008-9105-5.

[32] Gerson Graser, Bernd Schneider, Neil J Oldham, and Jonathan Gershenzon. "The Methionine Chain Elongation Pathway in the Biosynthesis of Glucosinolates in Eruca sativa (Brassicaceae)." In: *Archives of Biochemistry and Biophysics* 378.2 (2000), pp. 411–419. ISSN: 00039861. DOI: 10.1006/abbi.2000.1812. URL: http://linkinghub.elsevier.com/retrieve/pii/S0003986100918125.

[33] C. Douglas Grubb and Steffen Abel. "Glucosinolate metabolism and its control." In: *Trends in Plant Science* 11.2 (2006), pp. 89–100. ISSN: 13601385. DOI: 10.1016/j.tplants.2005.12.006.

[34] Stephane Guindon, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." In: *Systematic biology* 59.3 (2010), pp. 307–321.

[35] J.B.S. B S Haldane. *Enzymes*. London: Longmans, Green and Co. London, 1930. URL: http://www.ucl.ac.uk/Pharmacology/dc-bits/haldane-enzymes-1930-s.pdf.

[36] **Barbara Ann Halkier and Jonathan Gershenzon. "BIOLOGY AND BIOCHEMISTRY OF GLUCOSINOLATES." In: *Annual Review of Plant Biology* 57.1 (2006). Pp. 303–333. ISSN: 1543-5008. DOI: 10.1146/annurev.arplant.57.032905.105228. URL: http://www.annualreviews.org/doi/10.1146/annurev.arplant.57.032905.105228.**

*[36] An excellent review of glucosinolate research*

[37] Richard W Hamming. "Error detecting and error correcting codes." In: *Bell Labs Technical Journal* 29.2 (1950), pp. 147–160.

[38] Bjarne G Hansen, Daniel J Kliebenstein, and Barbara A Halkier. "Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis." In: *The Plant Journal* 50.5 (2007), pp. 902–910.

[39] Carsten Hørslev Hansen, Ute Wittstock, Carl Erik Olsen, Alastair J Hick, John A Pickett, and Barbara Ann Halkier. "Cytochrome P450 CYP79F1 from Arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates." In: *Journal of Biological Chemistry* 276.14 (2001), pp. 11078–11085.

[40] Yan He, Thomas P. Mawhinney, Mary L. Preuss, Amy C. Schroeder, Bing Chen, Linda Abraham, Joseph M. Jez, and Sixue Chen. "A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in Arabidopsis." In: *Plant Journal* 60.4 (2009), pp. 679–690. ISSN: 09607412. DOI: 10.1111/j.1365-313X.2009.03990.x.

[41] Reinhart Heinrich and Tom A. Rapoport. "A linear steady-state treatment of enzymatic chains. General properties, control and effector strength." In: *European journal of biochemistry / FEBS* 42.1 (1974), pp. 89–95. ISSN: 0014-2956. DOI: 10.1111/j.1432-1033.1974.tb03318.x. URL: http://doi.wiley.com/10.1111/j.1432-1033.1974.tb03318.x.

[42] Reinhart Heinrich and Stefan Schuster. *The Regulation of Cellular Systems*. Boston, MA: Springer US, 1996. ISBN: 978-1-4612-8492-5. DOI: 10.1007/978-1-4613-1161-4. URL: http://link.springer.com/10.1007/978-1-4613-1161-4.

[43] Victor Henri. *Lois generales de l'action des diastases.* Paris: Librairie Scientifique A. Hermann, 1903.

[44] Masami Yokota Hirai, Kenjiro Sugiyama, Yuji Sawada, Takayuki Tohge, Takeshi Obayashi, Akane Suzuki, Ryoichi Araki, Nozomu Sakurai, Hideyuki Suzuki, Koh Aoki, et al. "Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis." In: *Proceedings of the National Academy of Sciences* 104.15 (2007), pp. 6478–6483.

[45] L R Hogge, D W Reed, E W Underhill, and G W Haughn. "HPLC Separation of Glucosinolates from Leaves and Seeds of Arabidopsis thaliana and Their Identification Using Thermospray Liquid Chramatography/Mass Spectrometry*." In: *Journal of Chromatographic Science* 26 (1988). URL: http://blogs.ubc.ca/haughn/files/2012/04/Hogge-et-al.-1988-glucosinolate-analysis-in-Arabidopsis.pdf.

[46] John H. (John Henry) Holland and John H. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, 1992, p. 211. ISBN: 0262082136. URL: http://dl.acm.org/citation.cfm?id=531075.

[47] Mwafaq Ibdah, Ying-Tung Chen, Curtis G Wilkerson, and Eran Pichersky. "An aldehyde oxidase in developing seeds of Arabidopsis converts benzaldehyde to benzoic acid." In: *Plant Physiology* 150.1 (2009), pp. 416–423.

[48] H. Ishimoto, Y. Fukushi, T. Yoshida, and S. Tahara. "Rhizopus and Fusarium are Selected as Dominant Fungal Genera in Rhizospheres of Brassicaceae." In: *Journal of Chemical Ecology* 26.10 (2000), pp. 2387–2399. ISSN: 00980331. DOI: 10.1023/A:1005583012561. URL: http://link.springer.com/10.1023/A:1005583012561.

[49] Abhay Joshi and Bernhard O Palsson. "Metabolic dynamics in the human red cell: Part I—A comprehensive kinetic model." In: *Journal of theoretical biology* 141.4 (1989), pp. 515–528.

[50] H Kacser and J A Burns. "The control of flux." In: *Symposia of the Society for Experimental Biology* 27 (1973), pp. 65–104. ISSN: 0081-1386. URL: http://www.ncbi.nlm.nih.gov/pubmed/4148886.

[51] Shunro Kawakishi and Toshiyuki Kaneko. "Interaction of proteins with allyl isothiocyanate." In: *Journal of Agricultural and Food Chemistry* 35.1 (1987), pp. 85–88. ISSN: 0021-8561. DOI: 10.1021/jf00073a020. URL: http://pubs.acs.org/doi/abs/10.1021/jf00073a020.

[52] **D J Kliebenstein, J Kroymann, P Brown, A Figuth, D Pedersen, J Gershenzon, and T Mitchell-Olds. "Genetic control of natural variation in Arabidopsis glucosinolate accumulation." In: *Plant physiology* 126.2 (2001). Pp. 811–825. ISSN: 0032-0889. DOI: 10.1104/pp.126.2.811. URL: http://www.ncbi.nlm.nih.gov/pubmed/11402209http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC111171.**

*[52] Data on natural variation of GSL accumulation*

[53] Daniel J Kliebenstein. "A role for gene duplication and natural variation of gene expression in the evolution of metabolism." In: *PloS one* 3.3 (2008), e1838.

[54] Daniel J Kliebenstein, John C D'Auria, Aditi S Behere, Jae Hak Kim, Kevin L Gunderson, John N Breen, Grace Lee, Jonathan Gershenzon, Robert L Last, and Georg Jander. "Characterization of seed-specific benzoyloxyglucosinolate mutations in Arabidopsis thaliana." In: *The Plant Journal* 51.6 (2007), pp. 1062–1076.

[55] T. Knill, J. Schuster, M. Reichelt, J. Gershenzon, and S. Binder. "Arabidopsis Branched-Chain Aminotransferase 3 Functions in Both Amino Acid and Glucosinolate Biosynthesis." In: *PLANT PHYSIOLOGY* 146.3 (2008), pp. 1028–1039. ISSN: 0032-0889. DOI: 10.1104/pp.107.111609. URL: http://www.plantphysiol.org/cgi/doi/10.1104/pp.107.111609.

[56] Tanja Knill, Michael Reichelt, Christian Paetz, Jonathan Gershenzon, and Stefan Binder. "Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation." In: *Plant molecular biology* 71.3 (2009), pp. 227–39. ISSN: 1573-5028. DOI: 10.1007/s11103-009-9519-5. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2729411{\&}tool=pmcentrez{\&}rendertype=abstract.

[57] **Beate Knoke, Susanne Textor, Jonathan Gershenzon, and Stefan Schuster. "Mathematical modelling of aliphatic glucosinolate chain length distribution in Arabidopsis thaliana leaves." In: *Phytochemistry Reviews* 8.1 (2009). Pp. 39–51. ISSN: 15687767. DOI: 10.1007/s11101-008-9107-3.**

*[57] Skeleton model of chain-length distribution*

[58] Juergen Kroymann, Susanne Textor, Jim G. Tokuhisa, Kimberly L. Falk, Stefan Bartram, Jonathan Gershenzon, and Thomas Mitchell-Olds. "A Gene Controlling Variation in Arabidopsis Glucosinolate Composition Is Part of the Methionine Chain Elongation Pathway." In: *Plant Physiology* 127.3 (2001), pp. 1077–88. ISSN: 0032-0889. DOI: 10.1104/pp.010416. URL: http://www.plantphysiol.org/content/127/3/1077.longhttp://www.plantphysiol.org/content/plantphysiol/127/3/1077.full.pdfhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=129277{\&}tool=pmcentrez{\&}rendertype=abstract.

[59] Juergen Kroymann, Susanne Donnerhacke, Domenica Schnabelrauch, and Thomas Mitchell-Olds. "Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus." In: *Proceedings of the National Academy of Sciences of the United States of America* 100 Suppl.Supplement 2 (2003), pp. 14587–92. ISSN: 0027-8424. DOI: 10.1073/pnas.1734046100. URL: http://www.ncbi.nlm.nih.gov/pubmed/14506289http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC304123http://www.pnas.org/cgi/doi/10.1073/pnas.1734046100http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=304123{\&}tool=pmcentrez{\&}rendertype=.

[60] Joseph B Kruskal. "Nonmetric multidimensional scaling: a numerical method." In: *Psychometrika* 29.2 (1964), pp. 115–129.

[61] Virginia Lambrix, Michael Reichelt, Thomas Mitchell-Olds, Daniel J Kliebenstein, and Jonathan Gershenzon. "The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences Trichoplusia ni herbivory." In: *The Plant Cell* 13.12 (2001), pp. 2793–2807.

[62] Luca Lazzeri, Giovanna Curto, Onofrio Leoni, and Elisabetta Dallavalle. "Effects of Glucosinolates and Their Enzymatic Hydrolysis Products via Myrosinase on the Root-knot Nematode Meloidogyne incognita (Kofoid et White) Chitw." In: *Journal of Agricultural and Food Chemistry* 52.22 (2004), pp. 6703–6707. ISSN: 0021-8561. DOI: 10.1021/jf030776u. URL: http://www.ncbi.nlm.nih.gov/pubmed/15506804http://pubs.acs.org/doi/abs/10.1021/jf030776u.

[63] Chiu-Jong Lee and George S. Serif. "2-Amino-6-(methylthio)caproic acid, a methionine hemolog and precursor of progoitrin." In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 165.3 (1968), pp. 569–571. ISSN: 03044165. DOI: 10.1016/0304-4165(68)90246-8. URL: http://linkinghub.elsevier.com/retrieve/pii/0304416568902468.

[64] Jing Li, Bjarne Gram Hansen, James A Ober, Daniel J Kliebenstein, and Barbara Ann Halkier. "Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis." In: *Plant Physiology* 148.3 (2008), pp. 1721–1733.

[65] Qun Li, Sanford D. Eigenbrode, G. R. Stringam, and M. R. Thiagarajah. "Feeding and Growth of Plutella xylostella and Spodoptera eridania on Brassica juncea with Varying Glucosinolate Concentrations and Myrosinase Activities." In: *Journal of Chemical Ecology* 26.10 (2000), pp. 2401–2419. ISSN: 00980331. DOI: 10.1023/A:1005535129399. URL: http://link.springer.com/10.1023/A:1005535129399.

[66] Michael Lynch, Bruce Walsh, et al. *Genetics and analysis of quantitative traits*. Vol. 1. Sinauer Sunderland, MA, 1998.

[67] R Magrath, F Banot, M Morgner, I Parkin, A Sharpe, C Lister, C Deani, J Turnert, D Lydiate, and A Mithen. "Genetical Society of Great Britain Genetics of aliphatic glucosinolates. I. Side chain elongation in Brassica napus and A rabidopsis thaliana." In: *Heredity* 72 (1994), pp. 290–299. URL: https://www.jic.ac.uk/staff/caroline-dean/pdf{_}files/88{_}Magrath{_}R{_}et{_}al{_}1994{_}Heredity.pdf.

[68] M. Mari, O. Leoni, R. Iori, and T. Cembali. "Antifungal vapour-phase activity of allyl-isothiocyanate against Penicillium expansum on pears." In: *Plant Pathology* 51.2 (2002), pp. 231–236. ISSN: 0032-0862. DOI: 10.1046/j.1365-3059.2002.00667.x. URL: http://doi.wiley.com/10.1046/j.1365-3059.2002.00667.x.

[69] Anne-Claire M F Martines, Karen van Eunen, Dirk-Jan Reijngoud, and Barbara M Bakker. "The promiscuous enzyme medium-chain 3-keto-acyl-CoA thiolase triggers a vicious cycle in fatty-acid beta-oxidation." In: *PLoS computational biology* 13.4 (2017), e1005461. DOI: 10.1371/journal.pcbi.1005461. URL: http://www.ncbi.nlm.nih.gov/pubmed/28369071.

[70] I. Mewis, Ch. Ulrich, and W.H. Schnitzler. "The role of glucosinolates and their hydrolysis products in oviposition and host-plant finding by cabbage webworm, Hellula undalis." In: *Entomologia Experimentalis et Applicata* 105.2 (2002), pp. 129–139. ISSN: 0013-8703. DOI: 10.1046/j.1570-7458.2002.01041.x. URL: http://doi.wiley.com/10.1046/j.1570-7458.2002.01041.x.

[71] Elliot M Meyerowitz. "Prehistory and history of Arabidopsis research." In: *Plant physiology* 125.1 (2001), pp. 15–19.

[72] L. Michaelis and Miss Maud L. Menten. "The kinetics of invertin action: Translated by T.R.C. Boyde Submitted 4 February 1913." In: *FEBS Letters* 587.17 (2013), pp. 2712–2720. ISSN: 00145793. DOI: 10.1016/j.febslet.2013.07.015.

[73] Michael Dalgaard Mikkelsen, Peter Naur, and Barbara Ann Halkier. "Arabidopsis mutants in the C–S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis." In: *The Plant Journal* 37.5 (2004), pp. 770–777.

[74] Carol I. Miles, Marta L. del Campo, and J. Alan A. Renwick. "Behavioral and chemosensory responses to a host recognition cue by larvae of Pieris rapae." In: *Journal of Comparative Physiology A* 191.2 (2005), pp. 147–155. ISSN: 0340-7594. DOI: 10.1007/s00359-004-0580-x. URL: http://www.ncbi.nlm.nih.gov/pubmed/15711970http://link.springer.com/10.1007/s00359-004-0580-x.

[75] Ron Milo. "What is the total number of protein molecules per cell volume? A call to rethink some published values." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 35.12 (2013), pp. 1050–5. ISSN: 1521-1878. DOI: 10.1002/bies.201300066. URL: http://www.ncbi.nlm.nih.gov/pubmed/24114984http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3910158.

[76] R Mithen, J Clarke, C Lister, and C Dean. "Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in Arabidopsis thaliana." In: *Heredity* 74.2 (1995), p. 210.

[77] *NCBI Conserved Domain*. URL: https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?SEQUENCE=1063733424{\&}FULL (visited on 01/24/2018).

[78] Peter Naur, Bent Larsen Petersen, Michael Dalgaard Mikkelsen, Søren Bak, Hasse Rasmussen, Carl Erik Olsen, and Barbara Ann Halkier. "CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in Arabidopsis." In: *Plant physiology* 133.1 (2003), pp. 63–72.

[79] Nausicaa Noret, P Meerts, R Tolrà, C Poschenrieder, J Barceló, and J Escarre. "Palatability of Thlaspi caerulescens for snails: influence of zinc and glucosinolates." In: *New Phytologist* 165.3 (2005), pp. 763–772.

[80] AW Pearson, NM Greenwood, EJ Butler, and GR Fenwick. "Bochemical changes in layer and broiler chickens when fed on a high-glucosinolate rapeseed meal." In: *British poultry science* 24.3 (1983), pp. 417–427.

[81] Markus Piotrowski, Andreas Schemenewitz, Anna Lopukhina, Axel Müller, Tim Janowitz, Elmar W Weiler, and Claudia Oecking. "Desulfoglucosinolate sulfotransferases from Arabidopsis thaliana catalyze the final step in the biosynthesis of the glucosinolate core structure." In: *Journal of Biological Chemistry* 279.49 (2004), pp. 50717–50725.

[82] **A Radzicka and R Wolfenden. "A proficient enzyme." In: *Science* 267.5194 (1995). URL: http://science.sciencemag.org/content/267/5194/90.long.**

[83] L Rask, E Andreasson, B Ekbom, S Eriksson, B Pontoppidan, and J Meijer. "Myrosinase: gene family evolution and herbivore defense in Brassicaceae." In: *Plant molecular biology* 42.1 (2000), pp. 93–113. ISSN: 0167-4412. URL: http://www.ncbi.nlm.nih.gov/pubmed/10688132.

[84] A. Ratzka, H. Vogel, D. J. Kliebenstein, T. Mitchell-Olds, and J. Kroymann. "Disarming the mustard oil bomb." In: *Proceedings of the National Academy of Sciences* 99.17 (2002), pp. 11223–11228. ISSN: 0027-8424. DOI: 10.1073/pnas.172112899. URL: http://www.ncbi.nlm.nih.gov/pubmed/12161563http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC123237http://www.pnas.org/cgi/doi/10.1073/pnas.172112899.

[85] James Eric Rodman. "Population variation and hybridization in sea-rockets (Cakile, Cruciferae): seed glucosinolate characters." In: *American Journal of Botany* (1980), pp. 1145–1159.

[86] Julio C. Rojas. "Electrophysiological and Behavioral Responses of the Cabbage Moth to Plant Volatiles." In: *Journal of Chemical Ecology* 25.8 (1999), pp. 1867–1883. ISSN: 00980331. DOI: 10.1023/A:1020985917202. URL: http://link.springer.com/10.1023/A:1020985917202.

[87] M. Sarwar, J.A. Kirkegaard, P.T.W. Wong, and J.M. Desmarchelier. "Biofumigation potential of brassicas." In: *Plant and Soil* 201.1 (1998), pp. 103–112. ISSN: 0032079X. DOI: 10.1023/A:1004381129991. URL: http://link.springer.com/10.1023/A:1004381129991.

[88] Margret Sauter, Barbara Moffatt, Maye Chin Saechao, Rüdiger Hell, and Markus Wirtz. "Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis." In: *Biochemical Journal* 451.2 (2013), pp. 145–154.

[89] Yuji Sawada, Ayuko Kuwahara, Mutsumi Nagano, Tomoko Narisawa, Akane Sakata, Kazuki Saito, and Masami Yokota Hirai. "Omics-based approaches to methionine side chain elongation in Arabidopsis: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase." In: *Plant and Cell Physiology* 50.7 (2009), pp. 1181–1190.

[90] **Sascha Schäuble, Anne Kristin Stavrum, Pål Puntervoll, Stefan Schuster, and Ines Heiland. "Effect of substrate competition in kinetic models of metabolic networks." In: *FEBS Letters* 587.17 (2013). Pp. 2818–2824. ISSN: 00145793. DOI: 10.1016/j.febslet.2013.06.025. URL: http://**

*[82] Pioneer work about enzyme specificity*

*[90] Mathematical desription of substrate competition*

dx.doi.org/10.1016/j.febslet.2013.06.025http://ac.els-cdn.com/S0014579313004833/1-s2.0-S0014579313004833-main.pdf?{_}tid=bc0b879e-07fe-11e7-89d4-00000aacb35f{\&}acdnat=1489417811{_}ef1b58918c0014db12cd63456f56438dhttp://doi.wiley.com/10.1016/j.febslet.2.

[91] Klaus Schlaeppi, Natacha Bodenhausen, Antony Buchala, Felix Mauch, and Philippe Reymond. "The glutathione-deficient mutant pad2-1 accumulates lower amounts of glucosinolates and is more susceptible to the insect herbivore Spodoptera littoralis." In: *The Plant Journal* 55.5 (2008), pp. 774–786.

[92] J. Schuster, T. Knill, M. Reichelt, J. Gershenzon, and S. Binder. "BRANCHED-CHAIN AMINOTRANSFERASE4 Is Part of the Chain Elongation Pathway in the Biosynthesis of Methionine-Derived Glucosinolates in Arabidopsis." In: *THE PLANT CELL ONLINE* 18.10 (2006), pp. 2664–2679. ISSN: 1040-4651. DOI: 10.1105/tpc.105.039339. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.105.039339.

*[93] Details about shannon entropy*

[93] **C. E. Shannon. "A Mathematical Theory of Communication." In: *Bell System Technical Journal* 27.3 (1948). Pp. 379–423. ISSN: 00058580. DOI: 10.1002/j.1538-7305.1948.tb01338.x. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024.**

[94] B. J. Smith and J. A. Kirkegaard. "In vitro inhibition of soil microorganisms by 2-phenylethyl isothiocyanate." In: *Plant Pathology* 51.5 (2002), pp. 585–593. ISSN: 0032-0862. DOI: 10.1046/j.1365-3059.2002.00744.x. URL: http://doi.wiley.com/10.1046/j.1365-3059.2002.00744.x.

*[95] An excellent review of genes involved in GSL biosynthesis*

[95] **Ida E. Sønderby, Fernando Geu-Flores, and Barbara a. Halkier. "Biosynthesis of glucosinolates - gene discovery and beyond." In: *Trends in Plant Science* 15.5 (2010). Pp. 283–290. ISSN: 13601385. DOI: 10.1016/j.tplants.2010.02.005.**

[96] Susanne Textor, Stefan Bartram, Jürgen Kroymann, Kimberly L. Falk, Alastair Hick, John a. Pickett, and Jonathan Gershenzon. "Biosynthesis of methionine-derived glucosinolates in Arabidopsis thaliana : recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle." In: *Planta* 218.6 (2004), pp. 1026–1035. ISSN: 0032-0935. DOI: 10.1007/s00425-003-1184-3. URL: http://link.springer.com/10.1007/s00425-003-1184-3.

[97] Susanne Textor, Jan-Willem de Kraker, Bettina Hause, Jonathan Gershenzon, and James G Tokuhisa. "MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis." In: *Plant physiology* 144.1 (2007), pp. 60–71. ISSN: 0032-0889. DOI: 10.1104/pp.106.091579.

[98] K F Tierens, B P Thomma, M Brouwer, J Schmidt, K Kistner, A Porzel, B Mauch-Mani, B P Cammue, and W F Broekaert. "Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of Arabidopsis to microbial pathogens." In: *Plant physiology* 125.4 (2001), pp. 1688–99. ISSN: 0032-0889. DOI: 10.1104/PP.125.4.1688. URL: http://www.ncbi.nlm.nih.gov/pubmed/11299350http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC88826.

[99] Jim Tokuhisa, Jan-Willem de Kraker, Susanne Textor, and Jonathan Gershenzon. "Chapter two The biochemical and molecular origins of aliphatic glucosinolate diversity in Arabidopsis thaliana." In: 2004, pp. 19–38. DOI: 10.1016/S0079-9920(04)80003-0. URL: http://linkinghub.elsevier.com/retrieve/pii/S0079992004800030.

[100] Maria Traka and Richard Mithen. "Glucosinolates, isothiocyanates and human health." In: *Phytochemistry Reviews* 8.1 (2009), pp. 269–282. ISSN: 1568-7767. DOI: 10.1007/s11101-008-9103-7. URL: http://link.springer.com/10.1007/s11101-008-9103-7.

[101] B. Ulmer, C. Gillott, and M. Erlandson. "Feeding preferences, growth, and development of Mamestra configurata (Lepidoptera: Noctuidae) on Brassicaceae." In: *The Canadian Entomologist* 133.04 (2001), pp. 509–519. ISSN: 0008-347X. DOI: 10.4039/Ent133509-4. URL: http://www.journals.cambridge.org/abstract{_}S0008347X00006428.

[102] ***UniProtKB - Q8VX04 (MAM2_ARATH).* URL: http://www.uniprot.org/uniprot/Q8VX04 (visited on 01/18/2018).** *[102] MAM2 sequence*

[103] ***UniProtKB - Q9FG67 (MAM1_ARATH).* URL: http://www.uniprot.org/uniprot/Q9FG67 (visited on 01/18/2018).** *[103] MAM1 sequence*

[104] **Eberhard O Voit, Harald A Martens, and Stig W Omholt. "150 years of the mass action law." In: *PLoS computational biology* 11.1 (2015). e1004012. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004012. URL: http://www.ncbi.nlm.nih.gov/pubmed/25569257http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4288704.** *[104] An excellent review of the mass action law*

[105] A Werner and R Heinrich. "A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary" in vivo" state and of time dependent variations under blood preservation conditions." In: *Biomedica biochimica acta* 44.2 (1985), pp. 185–212.

[106] U. Wittstock, N. Agerbirk, E. J. Stauber, C. E. Olsen, M. Hippler, T. Mitchell-Olds, J. Gershenzon, and H. Vogel. "Successful herbivore attack due to metabolic diversion of a plant chemical defense." In: *Proceedings of the National Academy of Sciences* 101.14 (2004), pp. 4859–4864. ISSN: 0027-8424. DOI: 10.1073/pnas.0308007101. URL: http://www.ncbi.nlm.nih.gov/pubmed/15051878http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC387339http://www.pnas.org/cgi/doi/10.1073/pnas.0308007101.

# DECLARATION

I hereby declare that the thesis entitled "Mathematical models of glucosinolate metabolism in plants" has been composed solely by myself and it has not been submitted, in whole or in part, in any previous application for a degree.

*Düsseldorf, March 2018*

Suraj Sharma