Marchantia polymorpha as a testbed species for metabolic engineering



Anya Lindström Battle

Wadham College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2024

Declaration

I declare that all parts of this thesis are my own work. I would like to acknowledge that in Chapter 3, the data used to generate computational constraints on vacuolar organic acid storage in *Marchantia polymorpha* was collected by Corinna Hartinger (Department of Biology, University of Oxford, UK). In Chapter 6, the cloning used to generate constructs forming nuclear biomolecular condensates (plasmids GA2.1, GA3, pK7_GA2.1, and pK7_GA3) was carried out by Angus Barrett (Department of Biology, University of Oxford, UK). All images of nuclear biomolecular condensates were taken by Angus Barrett. In Chapter 5, the standard curves used for quantification of polar metabolites were generated and run on the GC-MS by Pedro Bota (Department of Biology, University of Oxford, UK).

This thesis has not been submitted, either partially or in full, for any degree at this university or another institution.

Anya Lindström Battle

Acknowledgements

First and foremost, I would like to thank my supervisor, Lee Sweetlove, for his guidance and encouragement over the last four years. My DPhil experience has been made all the more enjoyable by working with such a supportive supervisor. Thank you for giving me free reign but reigning me in when it was needed.

A number of people have provided technical know-how during the course of this project. I would like to thank Mark Fricker (for giving insights into all things microscopy-related), Laura Moody (for providing me with my *Marchantia* stock), and Radka Slovak and Susanna Streubel for sharing their wisdom regarding the transformation of *Marchantia*.

I would also like to thank Angus Barrett for his hard work during his Masters project, which laid the foundations for the synthetic biomolecular condensate work presented in this thesis. I am also grateful for the help and support from the other members of the Sweetlove group, namely Hilary Hunt, Sanu Shameer, Corinna Hartinger, Ed Smith, and András Sándor.

I would like to thank the people who keep the lab and the building facilities ticking over, namely Pedro Bota and Jack Bateman (for making sure the lab runs smoothly and always being accommodating), James Wisker and Rob Bryant (for keeping the building from falling down, quite literally sometimes), and Rebecca Horrell (for making sure my plants were always well taken care of).

I have enjoyed this DPhil journey, but it was by no means easy; during the hard periods it was invaluable to have somebody in the same situation with whom to share my pain. For this I am eternally indebted to the 'science girlies', namely Jacques Bouvier, Jochem Nielsen, and (especially) Lisa Zillig. Having a chat with you always made me feel better no matter what mistake I had made in the lab.

I am not exaggerating when I say I would not be where I am without the unconditional love and support of my family, and especially my parents and my brother. You are an anchor in who I am and in everything that I do. Mamma and dad, you have supported me throughout my life, and I will always be grateful for it. Tobias, thank you for knocking me down a peg or two whenever I need it. Please never stop.

Finally, I would like to thank my partner in life, Ricky. Thank you for always dragging me out of whatever mental hole I might have gotten myself into as a result of an experiment not working. Thank you for being so patient with me whenever I got home from the lab and was extremely hangry. Thank you for bringing out the best version of me.

Abstract

Establishing plant systems for economically viable production of high value biochemicals remains a formidable challenge due to the complexity and interconnectedness of central metabolism, resulting in unpredictable outcomes of metabolic engineering. *Marchantia polymorpha*, with its haploid-dominant lifecycle and ease of genetic manipulation, has emerged as a promising candidate for exploring plant metabolic engineering. This thesis investigates the potential of *M. polymorpha* as a testbed for metabolic engineering, focusing on the production of poly-3-hydroxybutyrate (PHB) as a case study. Challenges associated with PHB production such as low substrate availability and complex interactions with endogenous metabolism make it a good target pathway for developing metabolic engineering techniques. Computational modelling of *M. polymorpha* metabolism guided the identification of promising engineering targets in the cytosol, aiming to enhance PHB production while minimising negative impacts on plant growth. Additionally, the implementation of an optogenetic system enabled inducible PHB biosynthesis, demonstrating different effects on plant growth based on developmental stages. Furthermore, the targeting of PHB-producing enzymes to synthetic biomolecular condensates showed promise, although challenges remain in establishing such condensates in stable *Marchantia* lines. This study underscores the potential of *M. polymorpha* as a testbed species for advancing plant metabolic engineering, with implications for sustainable bioproduction systems.

Table of Contents

Declaration 3
Acknowledgements 4
Abstract 6
Table of Contents 7
Table of Figures and Tables 11
Abbreviations 14
1. Introduction 17
1.1. Thesis scope 17
1.2. Plant metabolic engineering 18
1.2.1 General challenges in plant metabolic engineering 18
1.2.2. Novel testbed species in plant metabolic engineering 22
1.3. The production of PHB as a case study 25
1.4. Spatiotemporal expression control to alleviate competition with endogenous metabolism 30
1.5. Compartmentalisation to control and increase metabolic flux 32
1.5.1. Probabilistic metabolite channelling occurs in large multi-enzyme clusters 34
1.5.2. Biomolecular condensates concentrate enzymes and affect their kinetics 36
1.5.3. The engineering of biomolecular condensates 40
1.6. Computational modelling can provide insights into metabolic networks 43
1.7. Thesis rationale and aims 49
2. Materials and methods 53
2.1. Computational modelling 53
2.1.1. Model construction 53
2.1.2. Estimation of non-growth associated maintenance (NGAM) costs 54
2.1.3. Setting up a diel model of *M. polymorpha* 54
2.1.4. Further model constraints 55
2.1.5. FBA and model analysis 57
2.1.6. Flux sampling 57
2.1.7. Data availability 59
2.2. Data analysis and statistics 59
2.3. Plant materials and growth 59
2.4. Molecular cloning and genetic construct fabrication 61

2.4.1. Genetic construct design 61
2.4.2. Genetic construct assembly 64
2.4.3. Transformation of *E. coli* 65
2.4.4. Isolation and analysis of plasmid DNA 66
2.4.5 Transformation of *A. tumefaciens* 67
2.5. Plant transformation 68
2.5.1. Transient transformation of *N. benthamiana* 68
2.5.2. Stable transformation of *M. polymorpha* 68
2.5.3. Extraction and purification of genomic DNA from *M. polymorpha* tissue 71
2.6. Plant image analysis 72
2.6.1. Macroscopic growth analysis of *M. polymorpha* 72
2.6.2. Tissue fixing and staining with Nile Blue A 72
2.6.3. Confocal laser scanning microscopy 73
2.6.4. Confocal image analysis 74
2.7. Protein purification and analysis 75
2.7.1. Detection and quantification of luciferase expression 75
2.7.2. Purification of overexpressed recombinant proteins from plant material 77
2.7.3. Analysis of proteins by sodium-dodecyl-sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) 78
2.7.4. Western blotting 78
2.8. Polymer and metabolite analysis 79
2.8.1. PHB extraction 79
2.8.2. Citramalate and general polar metabolite extraction 80
2.8.3. Gas chromatography-mass spectrometry (GC-MS) 81
3. Construction and analysis of a computational model to identify strategies for optimising the biosynthesis of PHB in *M. polymorpha* 83
3.1. Introduction 83
3.2. Model construction and testing 85
3.3. Modification of the baseline model for PHB synthesis 91
3.4. Optimal predicted paths to plastidial and cytosolic PHB production 95
3.5. Exploration of alternative engineering strategies using sampling of the flux solution space 99
3.5.1. Flux sampling reveals expected relationships between reactions 101
3.5.2. Alternative pathways predicted by flux sampling 103
3.5.3. Flux sampling conclusion 105
3.6. Investigation into the effect of sucrose on model conclusions 108

3.7. Discussion 110
3.7.1. Model analysis suggests that engineering the biosynthesis of PHB in the cytosol is a more promising strategy than plastidial production 110
3.7.2. Possible engineering targets for engineering cytosolic PHB production in *M. polymorpha* 112
3.7.3. Previous attempts at engineering acetyl-CoA availability and PHB synthesis ... 116
3.7.4. Limitations of the modelling approach 119
4. Development of *Marchantia polymorpha* as a metabolic engineering platform 123
4.1. Introduction 123
4.2. Using *M. polymorpha* as a testbed species for metabolic engineering 125
4.2.1. Transformation of *M. polymorpha* 125
4.2.2. *M. polymorpha* is insensitive to kanamycin 127
4.3. A fully functional optogenetic system for light-controllable transgene expression in *M. polymorpha* 128
4.3.1. The PULSE system 128
4.3.2. Characterisation of PULSE in *M. polymorpha* 130
4.4. Discussion 139
4.4.1. Challenges associated with working on an emerging testbed species 139
4.4.2. An evaluation of PULSE in *M. polymorpha* 139
4.4.3. Limitations of using PULSE in *Marchantia* 141
4.4.4. Outlook for using inducible promoters in the metabolic engineering of *M. polymorpha* 143
5. The production of PHB in *Marchantia polymorpha* 146
5.1. Introduction 146
5.2. The constitutive production of PHB in *M. polymorpha* 148
5.2.1. The design of a polycistronic construct for the constitutive expression of β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase 148
5.2.2. Characterisation of constitutive PHB production in *M. polymorpha* 150
5.3. Inducible control over PHB accumulation using an optogenetic system 154
5.3.1. The characterisation of *M. polymorpha* lines inducibly producing PHB 154
5.3.2. The dynamics of PHB accumulation in constitutive and inducible PHB-producing plants 160
5.4. Ancillary metabolic targets for increasing PHB accumulation in *M. polymorpha* 163
5.4.1. The design of a metabolic engineering strategy to increase cytosolic PHB accumulation in *M. polymorpha* 163
5.4.2. The effect of the overexpression of ACL and *At*tDT on the yield of PHB in constitutive PHB-producing lines 165

5.4.3. The effect of the overexpression of ACLY3KQ and *At*tDT on overall plant metabolite profile 169
5.5. Discussion 173
5.5.1. Negative effects associated with PHB production in *M. polymorpha* 173
5.5.2. Can inducible production alleviate phenotypic defects associated with PHB production? 174
5.5.3. The effects of computationally derived targets on the ability of *Marchantia* to support cytosolic PHB production 178
5.5.4. The effects of PHB production on the metabolism of *M. polymorpha* 181
6. Building synthetic biomolecular condensates in plants 185
6.1. Introduction 185
6.2. The design of a construct for building synthetic biomolecular condensates in plants 187
6.3. Characterisation of synthetic biomolecular condensates *in planta* 188
6.4. The targeting of enzymes to synthetic biomolecular condensates *in planta* 195
6.4.1. Citramalate synthase 195
6.4.2. The PHB pathway enzymes 200
6.4.3. Combination of synthetic biomolecular condensate-formation with other metabolic engineering tools 203
6.5. Discussion 206
6.5.1. Insights into the biophysical properties of synthetic biomolecular condensates in plants 206
6.5.2. The amenability of synthetic biomolecular condensates to metabolic engineering 209
7. Discussion 213
7.1. Summary of main findings 213
7.2. The production of PHB in *M. polymorpha* 216
7.3. Insights into metabolic bottlenecks to plant PHB production 218
7.4. Is *M. polymorpha* a good testbed species for metabolic engineering? 221
References 224
Supplementary Information 247

Table of Figures

Figure 1.1: PHB production in plants 30

Figure 3.1: Basic schematic of *M. polymorpha* baseline model showing selected inputs, outputs, and constraints 89

Figure 3.2: Major fluxes through the *M. polymorpha* baseline model with biomass accumulation as the objective 91

Figure 3.3: An illustration of the predicted optimal flux paths towards cytosolic (A) and plastidial (B) PHB production in *M. polymorpha* 97

Figure 3.4: Flux sampling and correlation analysis of the vacuolar citrate storage reaction 103

Figure 3.5: An illustration of main flux paths to PHB synthesis when the model is constrained based on flux sampling to carry flux through (A) nocturnal ACL, or (B) daytime peroxisomal malate synthase 107

Figure 3.6: The effect of increasing total sucrose import on selected fluxes, including on daytime and nocturnal sucrose import (A), and reactions previously shown to be involved in cytosolic PHB synthesis (B) 110

Figure 4.1: The PULSE optogenetic system and its activity under different light conditions 130

Figure 4.2: Characterisation of the PULSE optogenetic system in *M. polymorpha* 138

Figure 5.1: The characterisation of *M. polymorpha* lines constitutively producing PHB 153

Figure 5.2: The characterisation of *M. polymorpha* lines inducibly producing PHB 158

Figure 5.3: The dynamics of PHB accumulation in constitutive and inducible PHB-producing plants 162

Figure 5.4: The effect of the overexpression of ACLY3KQ or *At*ALB and *At*tDT on the yield of PHB in constitutive PHB-producing lines 168

Figure 5.5: The effect of the overexpression of ACLY3KQ and *At*tDT on overall plant metabolite profile 171

Figure 6.1: Characterisation of synthetic biomolecular condensates *in planta* 193

Figure 6.2: The localisation and activity of *Md*CMS in synthetic biomolecular condensates in plants 199

Figure 6.3: The localisation and activity of the three PHB-producing enzymes in synthetic biomolecular condensates in plants 202

Figure 6.4: Insights into citramalate and PHB production in biomolecular condensates by the co-infiltration of ancillary metabolic interventions 205

Figure S2.1: Superimposed light spectra of conditions used for growth of *M. polymorpha* in white light without (green) and with (red) supplementary far-red light 257

Figure S3.1: Pareto analysis of total PHB production versus biomass accumulation (top), and relative contributions to total PHB by night and daytime PHB production (bottom) 258

Figure S3.2: Flux sampling and correlation analysis of the cytosolic PHB accumulation reaction 259

Figure S3.3: Flux sampling and correlation analysis of the nocturnal ACL and peroxisomal malate synthase reactions 260

Figure S4.1: The number of publications per year making use of spore, thallus, or gemmae-based *M. polymorpha* transformation methods 261

Figure S5.1: An illustration of the different phenotypes obtained in screens of *M. polymorpha* expressing (A) constitutive PHB, (B) inducible PHB, and (C) ancillary intervention constructs 262

Figure S5.2: Impact on plant growth of induction conditions used in this study 263

Figure S5.3: Results of the metabolite screen 264

Figure S5.4: The effect of PHB accumulation on *M. polymorpha* thallus cuttings 265

Figure S6.1: Further characterisation of biomolecular condensates *in planta* 266

Figure S6.2: Alignment between *Md*CMS and MAM3 from *Arabidopsis* used to determine whether *Md*CMS contains a chloroplast transit peptide 268

Table of Tables

Table 1.1: Highest PHB yields achieved in the plastid and cytosol of selected plant species when constitutively producing the three-enzyme PHB pathway 29

Table 2.1: Reactions added to the baseline *M. polymorpha* model to allow for accumulation of PHB in the cytosol and in the plastid 56

Table 2.2: Gene parts used for the design of genetic constructs 63

Table 3.1: The flux to biomass accumulation in the baseline model, compared to the baseline model with starch storage or vacuolar malate/fumarate storage unconstrained 89

Table 3.2: Flux to biomass, flux to PHB, and sum of fluxes of the baseline model with and without cytosolic or plastidial PHB production 94

Table 3.3: Result of FVA between the baseline model and the plastidial or cytosolic PHB production models 99

Table 3.4: Possible engineering targets for the cytosolic production of PHB in *M. polymorpha* 115

Table S2.1: Constraints used for the generation of a stoichiometric model of *M. polymorpha* central metabolism 247

Table S2.3: Plasmids used in this study 249

Table S2.2: Primers used in this study 254

Table S3.1: Convergence diagnostics of 5,000 samples of the baseline and cytosolic PHB-producing models using the OPTGP algorithm 257

Abbreviations

ATP-citrate lyase ..... ACL

AUXIN RESPONSE FACTOR ..... ARF

Artificial intrinsically disordered peptide ..... A-IDP

Acetyl-CoA carboxylase ..... ACC

Human ATP-citrate lyase ..... ACLY

Aluminium activated malate transporter ..... ALMT

Cetyltrimethylammonium bromide ..... CTAB

Days post-infiltration ..... dpi

Design-build-test-learn ..... DBTL

4',6-diamidino-2-phenylindole ..... DAPI

Flux balance analysis ..... FBA

Flux variability analysis ..... FVA

Gas chromatography-mass spectrometry ..... GC-MS

Glycine-Serine-Glycine ..... GSG

Glyceraldehyde-3-phosphate dehydrogenase ..... GAPN

Gram fresh weight ..... gFW

Gram dry weight ..... gDW

Horseradish peroxidase ..... HRP

Helix-turn-helix ..... HTH

Heat shock protein ..... HSP

Isopentenyl pyrophosphate ..... IPP

Intrinsically disordered region ..... IDR

Intrinsically disordered peptide ..... IDP

Isocitrate dehydrogenase ..... IDH

Liquid-liquid phase separation ..... LLPS

Lysogeny broth ..... LB

Liquid nitrogen ..... LiN_2

Light-oxygen-voltage ..... LOV

Long-day ..... LD

Very long chain polyunsaturated fatty acid ..... LCPFA

Minimisation of metabolic adjustment ..... MOMA

N-methyl *N*-trimethylsilyltrifluoroacetamide ..... MSTFA

Citramalate synthase from *Malus domestica* ..... *Md*CMS

Methylthioalkylmalate synthase 3 ..... MAM3

Non-growth associated maintenance ..... NGAM

Sodium dichloroisocyanurate ..... NaDCC

Nuclear localisation signal ..... NLS

Oxaloacetate ..... OAA

Parsimonious flux balance analysis ..... pFBA

Poly-3-hydroxybutyrate ..... PHB

Pathogenesis-related ..... PR

Polyhydroxyalkanoates ..... PHA

Plant-usable light switch elements ..... PULSE

Phase separation coupled to percolation ..... PSCP

β-ketothiolase ..... PhbA

Acetoacetyl-CoA reductase ..... PhbB

PHB synthase ..... PhbC

Polymerase chain reaction ..... PCR

Principal component analysis ..... PCA

Passive lysis buffer ..... PLB

Pyruvate dehydrogenase ..... PDH

Phytochrome ..... Phy

Phytochrome interacting factor ..... PIF

Principal component 1 ..... PC1

Principal component 2 ..... PC2

Sodium-dodecyl-sulphate-polyacrylamide gel electrophoresis ..... SDS-PAGE

Stress granule ..... SG

Standard error of the mean ..... SEM

Single ion monitoring ..... SIM

tert-butyldimethylsilane ..... tBDMS

Long-day white light conditions ..... W

Long-day white light conditions with supplemental far-red ..... W+FR

Wild-type ..... WT

1. Introduction

1.1. Thesis scope

Metabolic engineering holds substantial promise to generate modified organisms capable of producing high-value molecules in a renewable fashion and stands to play a major role in the 'green economy' that we must transition to in the coming decades. Metabolic engineering involves the modification of metabolic pathways to redirect metabolism towards maximizing a product of interest. Because metabolism forms an interconnected network, metabolic engineering involves a consideration of the metabolic system in its entirety.

Although metabolic engineering has been extensively implemented in microbes, the metabolic engineering of plants to produce valuable compounds offers certain advantages. The main obstacle to commercially-viable microbial production strategies is their cost, up to 60% of which comes from the carbon substrate used in fermentation (*1*). As autotrophic organisms, plants are able to use sunlight and atmospheric CO_2 for growth, raising the possibility of carbon neutral production (*2*). There is also increasing interest in developing plants which accumulate high-value compounds alongside the biomass required for biofuel production as a means of improving the economics of biorefineries and making biofuel cost-competitive (*3*).

Progress in the metabolic engineering of plants has lagged behind advances made in other systems; although complex genetic circuits have been built in bacteria, yeast, and mammalian cells, the first paper describing the design and assembly of simple

genetic circuits in plants was published only in 2022 (*4*). Part of the reason for this is the long generation time of plants and the laborious nature of plant transformation. For this reason, my project aims to test whether the bryophyte *Marchantia polymorpha* can be leveraged to accelerate metabolic engineering in plants. The prolonged haploid phase of *M. polymorpha* reduces the time taken to obtain homozygous mutants to 4-6 weeks, compared to 5-6 months for *Arabidopsis thaliana* and 12-18 months for many crop plants (*5*).

As well as long generation times, the engineering of plants suffers from many of the same obstacles to achieving high product titre as microbial engineering such as product- or intermediate-toxicity, competing endogenous reactions, or negative effects of engineering interventions on, for instance, biomass accumulation (*6*). The effect of a metabolic modification can often be unexpected and unintuitive meaning several rounds of intervention and testing are required to optimise the engineering intervention, further compounding issues relating to generation times. Using *M. polymorpha* and the production of poly-3-hydroxybutyrate (PHB) as a case study, I have developed a set of tools and techniques that can be used to explore and overcome some of these challenges. These advances will help establish *M. polymorpha* as a promising testbed species for metabolic engineering and develop plant metabolic engineering in general.

1.2. Plant metabolic engineering

1.2.1 General challenges in plant metabolic engineering

Metabolic networks are complex and interconnected. This is especially true of central

or primary metabolism which typically carries high flux and is responsible for the transformation of input nutrients into energy and the precursor metabolites necessary for biomass accumulation (*7*). Central metabolism is organized into a highly conserved 'bow tie' structure, whereby all inputs are converted into one of 12 precursor metabolites. Each of these is then involved in a large number of biosynthetic reactions in order to produce all cellular components required for the cell (*8*). Cells have evolved multiple layers of regulation such that the production of cellular components is balanced with energy production and consumption. Although many attempts at rational engineering of central plant metabolism have been made, there exist few examples where the engineering intervention has had the desired outcome (*9*). The recalcitrance of central metabolism to engineering is due to its tight and multilayered regulation as well as its interconnectedness. As such even if a change in flux is achieved in a particular pathway this is likely to have unforeseen circumstances elsewhere in the network.

This can be illustrated by a study attempting to ascertain the function of the mitochondrial acetyl-CoA pool. The authors generated transgenic tobacco plants expressing an acetyl-CoA hydrolase from yeast, which was targeted to the mitochondria using a mitochondrial targeting peptide from a potato processing peptidase. The introduced enzyme was correctly targeted and transformants displayed reductions in citrate and malate levels as well as increased acetate content, as would be expected given higher mitochondrial acetyl-CoA hydrolase activity. Surprisingly, however, the transgenic plants also exhibited significantly increased total acetyl-CoA levels, an effect at least partly attributed to the activity of acetyl-CoA

synthetase which was up to 3-fold higher than in wild-type (WT). In addition, transgenic plants displayed dramatically reduced growth and development, as well as leaf chlorosis. This severe growth phenotype was accompanied by a general negative effect on photosynthesis, including reduced levels of photosynthetic pigments as well as reduced CO_2 assimilation, transpiration rate, and stomatal conductance. Taken together, this example illustrates the often unexpected effects of engineering interventions on central metabolism, and that disruption of a central metabolite like acetyl-CoA can have severe metabolic and developmental consequences (*10*).

Another example of the interconnectedness of metabolism can be illustrated using substrate availability. Limited substrate availability for an engineered pathway has been shown in several instances to affect the outcome of the metabolic intervention (*11-13*). There are several ways in which adequate substrate availability may be achieved *in vivo*. One might: (a) target enzymes to parts of the cell with known higher concentrations of the substrate; (b) increase flux towards substrate production; (c) reduce flux through pathways known to consume the substrate itself (*5, 14, 15*). There is no obvious way of choosing between these strategies and in fact the best outcome may be achieved by a combination of several of them. As an example, consider one of the most successful targets for metabolic engineering, namely the anti-malarial drug artemisinin from *Artemisia annua*. By directing the committed step of the artemisinin pathway to mitochondria, known to contain higher levels of the precursor isopentenyl pyrophosphate (IPP), an 8-fold increase in artemisinin production over cytosolically driven production was achieved. In addition, the authors co-expressed a deregulated version of 3-hydroxy-3-methylglutaryl-coenzyme A reductase from yeast to increase

supply of precursor from the cytosolic mevalonate pathway (*16*). Even higher yields, approaching industrially relevant levels, were obtained when the entire mevalonate pathway was co-expressed in the chloroplast with the artemisinin biosynthetic pathway genes, to ensure adequate precursor availability given incomplete knowledge regarding the transport of mevalonate pathway metabolites into the plastid (*17*).

Whilst efficient at increasing yield, this latest strategy also had unwanted and detrimental effects on plant growth and development. The transformants grew at a slower rate than wild-type plants, were stunted and failed to induce proper root fibres when grown *in vitro* (*18*). This example shows that the manipulation of secondary metabolism can impact primary metabolic processes, despite the assumption that primary metabolic pathways are capable of supporting an excess of secondary metabolite synthesis due to the maintenance of sufficient substrate level pools (*14*).

In some instances, the engineering intervention can have no effect at all; this is because of a complex network of regulatory mechanisms which can act to negate or completely bypass the intervention. What is more, the exact nature of these control systems is still incompletely understood, but are thought to act in a hierarchical manner from transcription through translation to enzyme activity, whilst feedback mechanisms further complicate the picture (*19*). The advent of 'omics technologies and the integration of this information into computational models has helped gain a fuller understanding of these regulatory networks (*20-22*), but even in the best studied organisms such as *Escherichia coli* or *Saccharomyces cerevisiae* this knowledge is incomplete. In eukaryotes, gene regulation is even more extensive and multilayered, acting for instance through non-coding RNAs and epigenetics. Plants are particularly

reliant on epigenetic regulation because of their development and lifestyle. All the major epigenetic mechanisms characterised in eukaryotes are known to occur in plants, reflecting the fact that plants continuously produce new organs from their meristems as opposed to being specified during embryonic development. In addition, environmental conditions can shape the postembryonic development of plants allowing for phenotypic plasticity, an important attribute given plants are sessile and thereby required to cope with changing environmental conditions as well as biotic stress (*23, 24*).

1.2.2. Novel testbed species in plant metabolic engineering

The complexity, unpredictability, and tight regulation of plant metabolism present a challenge to the generation of rationally engineered plants. In microbial engineering, in which these challenges also occur, the 'design-build-test-learn' (DBTL) cycle refers to the parts of the iterative process often used to overcome them; a biological system is designed and translated into DNA parts before being implemented in a chassis organism, the engineered pathway is examined to determine if and why it functions sub-optimally, and these lessons are then applied to the next round of engineering (*8*). This cycle may be repeated multiple times until the desired outcome is achieved. This process is complicated in plants because of much longer generation times and a laborious process for obtaining transgenics, restricting the number of transformation cycles that are realistically feasible.

To iteratively test metabolic engineering changes, a plant that is easy to transform, fast, and an adequate model for potential production hosts (a so-called 'testbed' plant) is

needed. For *A. thaliana*, the most commonly used model species in plant science, it takes 5-6 months to generate a stable transgenic line (*5*). Although not slow by plant standards (consider the 7-10 year generation cycle of poplar, a model species used to study woody plants (*25*)), this is still significant compared to the mere weeks it takes to design, transform, and test a bacterial strain carrying a novel metabolic intervention. In terms of speed, transient transformation of the leaves of *Nicotiana* spp. with *Agrobacterium tumefaciens* most approaches microbial engineering. The most commonly used species is *Nicotiana benthamiana*. Results may be obtained as soon as several days after infiltration and very high yields of up to 1.5 g kg^{-1} fresh weight of GFP have been achieved (*26*). In addition, combinatorial infiltration with *Agrobacterium* strains carrying more than 10 different unique expression constructs can be routinely carried out. This allows for rapid testing of multigene biosynthetic pathways, as well as the combination of different variants such as different enzymes or promoter strengths (*27*). This approach has been particularly successful in the design and testing of heterologous pathways for secondary metabolite synthesis, such as the 16-gene pathway for (−)-deoxypodophyllotoxin which accumulated to 4.3 mg g^{-1} dry weight (*28*). Transient expression in *Nicotiana tabacum* has been used on a commercial scale, for example to rapidly produce antibodies for the treatment of Ebola (*29*).

However, efficient transient expression using *A. tumefaciens* has only been achieved for *Nicotiana* spp. Transient expression methods have been established for *A. thaliana* as well as seven other plant species (*30*), but remain highly unreliable. In addition, transient expression is short-lasting, and in most cases will decrease to negligible

levels within 8 days post-infiltration (*31*). This makes it unsuitable for studying long-term expression or more systemic effects such as on biomass accumulation. Lastly, agroinfiltration is known to elicit an immune response in the agroinfiltrated tissue, leading to induction of pathogenesis-related (PR) genes, the accumulation of extracellular PR proteins, and the down-regulation of photosynthesis-associated genes, raising concerns about the suitability of this platform to study non-diseased metabolism (*32*). Nevertheless, the use of transient expression is still a very powerful tool to speed up the DBTL cycle by allowing for the testing of constructs prior to the inevitably longer process of generating stable transgenic plants.

Bryophytes have also been suggested as potential testbed organisms for plant metabolic engineering. Bryophytes are the closest living relatives to the first land plants, and are comprised of the Marchantiophyta (liverworts), Bryophyta (mosses) and Anthocerotophyta (hornworts). In contrast to most other land plants, the haploid gametophyte generation is the dominant stage in bryophytes, allowing for the faster generation of homozygous mutants. The species for which engineering techniques are the most developed are the moss *Physcomitrium patens* and the liverwort *M. polymorpha*. Both these species can be easily maintained in axenic culture using standardized cultivation conditions, and possess a unique regeneration capacity from fragments or single cells allowing for easy and rapid propagation (*33*). Importantly, efficient transformation methods for genetic engineering using *A. tumefaciens* or polyethylene glycol exist for both species (*34-36*). More recently, CRISPR (*37, 38*) and standard synthetic biology parts (*39-42*) have also been developed for both *P. patens* and *M. polymorpha*. As a result, *P. patens* has been engineered to produce artemisinin

to levels equivalent to those in *A. annua* (a result which has been hard to achieve in other production hosts, including microbes) by the heterologous expression of five genes (*43*), and is also used in the perfume industry to produce sesquiterpenoids such as patchoulol (*44*). In *M. polymorpha*, the natural accumulation of arachidonic acid has been taken advantage of to produce prostaglandins by the expression of a cyclooxygenase gene from *Gracilaria vermiculophylla*, representing the first system for the bioproduction of prostaglandins in a plant (*45*).

1.3. The production of PHB as a case study

Of the many high-value molecules that have been considered for production by plants, bioplastic has perhaps the greatest economic and environmental benefits (*3*). Conventional plastics are produced non-renewably from petroleum and, as nondegradable products, pose issues both to the environment and human health. Biodegradable polyesters, and particularly polyhydroxyalkanoates (PHAs), are emerging as a way of mitigating these effects, and may become critical given the increasing dependence on and demand for plastics (*46, 47*). PHAs are synthesized by most genera of eubacteria for carbon storage under stress conditions (*48*). The simplest form of PHA is PHB, a precursor of propylene. PHB is an end-product of acetyl-CoA anabolic metabolism and its biosynthesis involves the enzymes β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase (*46*) (Figure 1.1A). In the first step, β-ketothiolase condenses two molecules of acetyl-CoA to acetoacetyl-CoA. In the second step acetoacetyl-CoA reductase catalyses the formation of 3-hydroxybutyryl-CoA with the cost of NADPH oxidation to $NADP^+$. Finally, PHB synthase polymerizes 3-hydroxybutyryl-CoA monomers into PHB (*49*). Alternatively, β-

ketothiolase can be replaced by *NphT7*-encoded acetoacetyl-CoA synthase from *Streptomyces* sp. which converts acetyl-CoA and malonyl-CoA to acetoacetyl-CoA (*50*). Engineering of bacteria for PHB production has been relatively successful, but the commercial viability of this strategy is limited by the factors already discussed.

The engineering of PHB production in plants has been widely attempted but suffers from many of the challenges facing plant metabolic engineering already explained, and thus presents itself as a good case study for overcoming them. The engineering of plants for bioplastic production has been an active research field for the best part of three decades, ever since the report of PHB production in *A. thaliana* in 1992 by the constitutive expression of *Ralstonia eutropha phbB* and *phbC* genes, which took advantage of endogenous β-ketothiolase activity (*51*). To date the highest yields of PHB *in planta* have also been reported in *A. thaliana*, with a report of up to 40% PHB per dry weight in leaf tissue published in 2000 (*52*). However, this resulted in severely stunted plants with extensive chlorosis and loss of fertility (Figure 1.1B). In fact, to my knowledge all cases in which PHB has been produced to high levels has been accompanied by detrimental effects on plant health (Table 1.1). Recently, a large language model was used to generate the top 100 most important questions facing plant science, one of which was 'How can we use plant-based materials to develop sustainable plastics?' (*53*), showing that plant-based bioplastic production remains an important research area.

Organelle-specific PHB production has been explored, with pathways introduced into the cytosol, plastids, peroxisomes, and mitochondria of various plant systems (*54*). This work will focus on cytosolic and plastidial PHB production, as will be explained in

Chapter 3. The highest yields of PHB have been achieved when PHB production has been localized to the plastid (*55*) (Table 1.1); this probably reflects access to higher levels of acetyl-CoA due to the role of plastids in fatty acid synthesis (*56*). The previously mentioned yield of 40% per unit leaf dry weight was achieved by targeting the three enzymes of PHB synthesis to the chloroplast using a pea chloroplast transit peptide (*56*). The enzymes have also been produced directly in the chloroplast via plastome transformation of tobacco, leading to an accumulation of 18.8% PHB per unit dry weight of the leaves, more than 60-fold that achieved with nucleus-encoded genes in the same organism (*57*). Production of PHB in the chloroplasts of switchgrass leaves has led to accumulation of PHB to 7.7% per unit leaf dry weight (*58*). In sugarcane, an average of 6.3% per unit dry weight of total leaf biomass was achieved when PHB was produced in the chloroplasts of mesophyll cells (*59*).

The strong negative growth phenotype associated with PHB production in plants has resulted in low and non-cost-competitive product titres. The accepted estimate is that PHB must accumulate to 10-12.5% per unit whole-plant dry weight for the process to be cost-competitive with petroleum-derived plastics (*46, 47*). Yields are often only reported for specific organs (e.g. leaves) which complicates numerical comparisons, but it is clear that PHB levels achieved in any viable production plant are still far off this value.

Several reasons have been put forward for these phenotypic effects and depend on the location of PHB production both on a subcellular and tissue level. One hypothesis is that there is competition between PHB production and other biomass-accumulation pathways for acetyl-CoA, a key metabolite in the central metabolism of all organisms

connecting catabolic and anabolic metabolism (*60*). In eukaryotes, acetyl-CoA cannot be transported between the cytosol and other organelles (*46, 60*). As a result the overexpression of the cytosolic enzyme ATP citrate lyase (ACL), which generates acetyl-CoA from citrate and CoA, is able to partly compensate for cytosolic PHB synthesis-associated growth defects in *A. thaliana* (*61*). This competition between biomass accumulation and PHB production is generalizable to any system in which PHB is produced, as shown by the modelling of PHB production in *Paracoccus denitrificans*. In this study PHB production was shown to occur mostly in the absence of growth; during the exponential growth phase acetyl-CoA was shuttled into the TCA cycle and was no longer available for β-ketothiolase (*62*).

In the engineering of sugarcane and switchgrass for plastidial PHB production, adverse phenotypes such as stunting, chlorosis and reduced biomass were observed with yields above approximately 1.5% per unit dry weight (*49*). These are likely due to the excess use of chloroplast acetyl-CoA for PHB biosynthesis, reducing its availability for fatty acid and isoprenoid biosynthesis. In addition, it has been suggested that the physical presence of PHB granules can shade the thylakoids, leading to reduced photosynthesis and ATP starvation (*48, 63*). It is also clear that the accumulation of PHB granules in the plastid leads to severely abnormal plastid morphology (Figure 1.1C) (*52, 56*).

As such, PHB production in plants presents challenges in terms of substrate availability, low product titre, and complex interactions with endogenous metabolism, issues which are generally applicable to plant metabolic engineering. For these reasons PHB production was selected in this project as a good target pathway for the

development of plant metabolic engineering.

Table 1.1: Highest PHB yields achieved in the plastid and cytosol of selected plant species when constitutively producing the three-enzyme PHB pathway.

Species	Location	Yield (% PHB per unit dry weight, unless otherwise indicated)	Adverse phenotypic effects observed	Ref.
N. tabacum	Plastid	18.8 (leaves); 8.8 (whole plant)	Delayed flowering, slower growth, delayed development	(*57*)
	Cytosol	0.09 (leaves)	None reported	(*64*)
A. thaliana	Plastid	40 (leaves)	Stunted growth, loss of fertility, chlorosis in leaves	(*52*)
	Cytosol	0.05 (fresh weight, leaves)	Severe growth reduction, minimal seed production	(*65*)
Saccharum officinarum	Plastid	6.3 (leaves)	Stunting, slow growth, chlorosis	(*59*)
	Cytosol	Trace levels (leaves)	None reported	(*66*)
Panicum virgatum	Plastid	7.7 (leaves); 1.7 (whole tiller)	Unknown	(*58*)

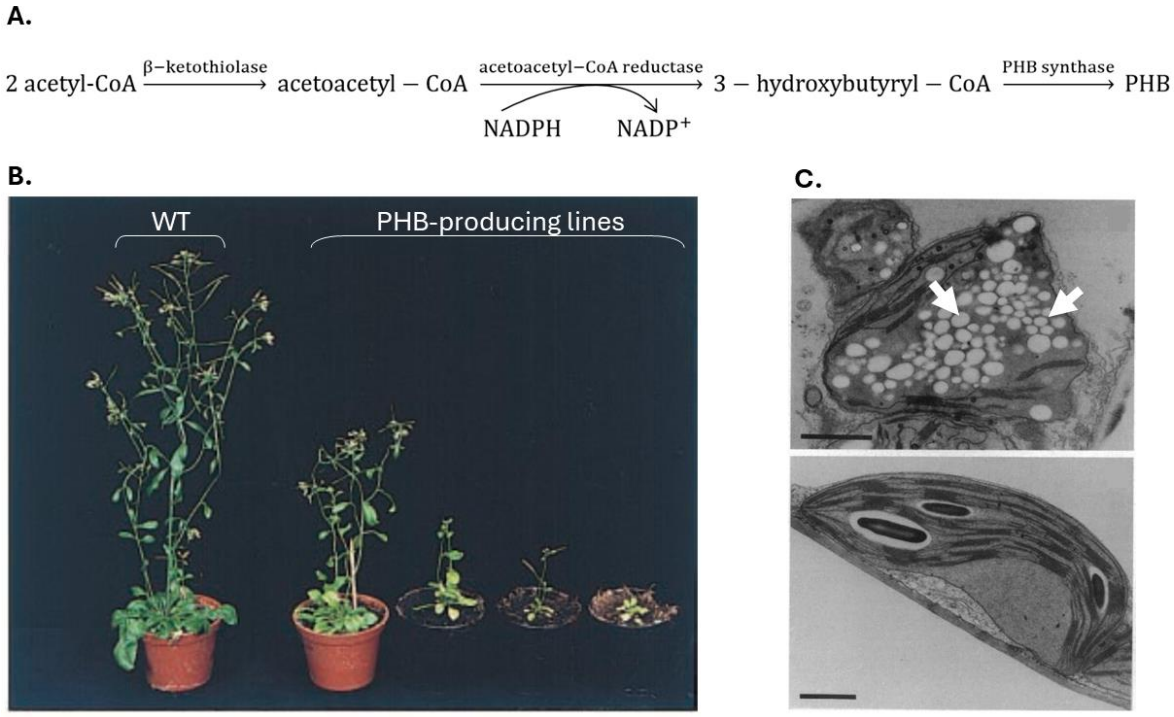


Figure 1.1: PHB production in plants. (A) The PHB production pathway from *R. eutropha* commonly used to produce PHB in plants. (B) 7-week-old plastidial PHB-producing *Arabidopsis* plants show severe growth retardation compared to wild-type. Adapted from (52). (C) Transmission electron micrographs reveal that the accumulation of 0.2 – 0.7 µm PHB granules within the plastids of *Arabidopsis* (top, indicated by arrows) leads to abnormal plastid morphology compared to wild-type (bottom). Scale bar is 1 µm. Adapted from (56).

1.4. Spatiotemporal expression control to alleviate competition with endogenous metabolism

The constitutive production of PHB causes a marked negative growth phenotype, probably at least partly due to competition between introduced and endogenous metabolism for acetyl-CoA. Inducible systems have been used to overcome deleterious or lethal effects of product accumulation or metabolic disruption in plant systems. For example, in transient protein production in *N. benthamiana*, the high

accumulation levels of some proteins can be toxic to the plant. The ethanol-inducible *AlcA* promoter and the transcription factor AlcR from *Aspergillus nidulans* have been extensively used to drive high protein expression prior to harvest (*67*). Also in *N. benthamiana*, a heat-shock inducible promoter has been used to produce bioactive casbene-derived terpenoids (*68*), whose constitutive production had been accompanied by delayed growth and fitness penalties (*69*).

In a similar way an inducible promoter could be used to avoid the competition between PHB production and endogenous metabolism by, for instance, inducing expression at a specific point in the plant lifecycle (*46-48*). For example, the ethanol-inducible system was used for plastidial PHB production in tobacco and resolved growth deficiency and male sterility phenotypes previously observed; however it resulted in yield of PHB to only 0.2% per unit whole-plant dry weight (*70*). Other examples include the use of salicylic acid (*71*) or ecdysone (*72*) to induce the expression of genes encoding PHB-synthesis enzymes. Similarly, enzyme inhibitors have been used to increase cytosolic PHB production in mature tobacco plants. A tobacco line that constitutively produced PHB in the cytosol was treated with the acetyl-CoA carboxylase inhibitor D-(+)-Quizalofop-ethyl which increased PHB accumulation 2-fold (albeit to 0.09% per unit leaf dry weight), although subsequent effects on plant growth were not quantified (*64*). However, most of the inducible promoters used to date are leaky which limits their ability to remove phenotypic defects (*46*). Additionally, they do not drive sufficiently high levels of gene expression to lead to economically relevant amount of PHB accumulation.

In contrast to these chemically inducible systems, optogenetic systems (that is gene

expression regulated by light) can offer spatiotemporal, quantitative, controllable, fully orthogonal, and reversible control over metabolic processes (*73*). Although optogenetics has been extensively used in prokaryotic and mammalian systems, the requirement of plants for light to grow complicates the implementation of optogenetic approaches in these organisms. This is because ambient light leads to undesired optogenetic system activation, and is incompatible with most previously published optogenetic systems (*74*).

Recently, the 'plant usable light-switch elements' (PULSE) system has been developed which allows for the control of plant gene expression under ambient light (*73*). In this system, gene expression is only activated by monochromatic red light; under white light or in darkness it remains inactive. Importantly, this system works in a transgenic *A. thaliana* plant, and thus shows great promise as a tool in plant biotechnology. The PULSE system relies on two interacting engineered modules, one of which represses gene expression in the presence of blue light (B_{off}; engineered from the EL222 photoreceptor) and the other which activates gene expression with red light (R_{on}; engineered from the phytochrome B – PIF6 optoswitch). This system has a high dynamic range of expression of up to 400-fold induction in protoplasts, and 21-fold in stably transformed *A. thaliana*, with variable levels of background expression.

1.5. Compartmentalisation to control and increase metabolic flux

Although optogenetic techniques can be used to provide temporal control over gene expression, it would be desirable to combine this with further control over competition

for acetyl-CoA from for instance, biomass-producing reactions. Compartmentalisation is a well-characterised method used by organisms from across the tree of life to control metabolism. The most well-studied compartments are organelles, but more recently compartmentalisation into biomolecular condensates has emerged as an important way in which cells divide up their intracellular environment. The idea that cells organize and concentrate components into nonmembrane-bound organelles has emerged in the last 20 years; protein and/or nucleic acid condensates such as nucleoli, Cajal bodies, or pyrenoids have been discovered throughout the tree of life (*75*). A recent large-scale localization study of chloroplast proteins revealed a large number of novel punctate structures composed of metabolic enzymes, and proposed to play a role in the spatial organization of biosynthetic reactions such as L-serine biosynthesis or the branch point between fatty acid synthesis and the production of glutamate (*76*).

Biologically, these membraneless compartments have been suggested to carry out a range of functions linked to control over metabolic flux, including spatiotemporal compartmentalisation of specific biochemistry, tuning reactions by providing specific microenvironments, and dynamically responding to stress (*75, 77*). Recent work has demonstrated in detail the kinetic benefits that can be imparted on enzymes localized to such condensates (*78*), suggesting another reason why these structures are so ubiquitous in biology. Control over metabolic flux, for example at metabolic branch points, and increased metabolic flux through a pathway are important tools in the metabolic engineering toolbox, suggesting a potential role for biomolecular condensates in the engineering of biological systems.

1.5.1. Probabilistic metabolite channelling occurs in large multi-enzyme clusters

A potential secondary benefit of localising enzymes to a biomolecular condensate is that pathway flux may increase due to the high concentration of enzymes and the emergence of metabolite channelling. Evidence has accumulated over the last 40 years to suggest that the view of metabolic systems as homogeneously distributed enzymes and freely diffusing metabolites does not hold true for metabolism *in vivo*. Instead, enzymes, often members of recognized metabolic pathways, dynamically coalesce in response to biochemical stimuli. Metabolites may be channelled between sequential enzymes of a metabolic pathway without equilibration with the bulk aqueous phase of the cell, a phenomenon known as metabolite channelling. Enzyme systems supporting metabolite channelling are known as 'metabolons' and have been observed in a wide variety of pathways from glycolysis to polyketide biosynthesis and have emerged as an important feature of metabolism across multiple types of organisms (*79*).

These enzyme assemblies have been shown to broadly fall into two categories. In the first, single complexes of paired enzymes direct the diffusion of the channelled intermediate, for instance by co-evolved complementary structural and charge surface features. An analysis of the modelled malate dehydrogenase-citrate synthase-aconitase complex showed positively charged residues localized to a groove which connects the malate dehydrogenase and citrate synthase active sites. Electrostatic retention of the negative oxaloacetate intermediate within this groove would promote efficient transfer between active sites (*79*). Another example of this type of enzyme

assembly is the bifunctional tryptophan synthase, which uses a 2.5 nm-long hydrophobic tunnel to transport indole from one active site to another (*80*).

The second type of enzyme assembly is formed from large clusters of multiple copies of enzymes from a pathway, which may or may not be ordered. In these clusters, specific pairwise interaction between sequential enzymes is no longer required for substrate channelling, and the channelling effect is probabilistic: when a product diffuses away from the active site of one enzyme in the cluster, the probability of encountering the next enzyme in the pathway is much higher than the metabolite diffusing into the bulk aqueous solvent of the cell (*79, 81, 82*). Such large multi-enzyme complexes have been seen in nature. Examples include a cAMP microdomain formed of oligomerising adenyl cyclases, phosphodiesterases, and A-kinase anchoring proteins, or the post-synaptic density in neurons, a 300 nm-diameter high-order complex which contains an estimated 10^4 proteins (*81*). In plant metabolism, the TCA cycle and the glycolytic pathway have been suggested to mediate channelling by the formation of supramolecular complexes (*83*).

Both these types of enzyme assemblies have in common their ability to bypass branch points in the metabolic network, and thereby change the structure of the network and regulate metabolic flux, and to be able to sequester labile and toxic intermediates from the bulk aqueous body of the cell. Probabilistic channelling would be easier to exploit in synthetic biology than trying to engineer structural compatibility between enzymes which have not co-evolved (*82*). In addition, probabilistic channelling theoretically has several attributes which makes it desirable as a tool in metabolic engineering, not least an enhanced overall reaction rate. In large enzyme clusters the local enzyme

concentration is increased, leading to a kinetic benefit for enzyme systems reaching substrate saturation (*79*). These effects can be significant, with suggestions of 100-fold flux increases for a three-step pathway compared to freely diffusing enzymes (*84*). In enzyme assemblies formed of pairwise interactions, however, direct channelling is not expected to increase the steady-state reaction rate. This is because the vast majority of non-channelled reactions are not diffusion limited; in fact, the time taken for metabolites to diffuse between enzymes has been estimated to be 1,000-10,000 times faster than the average reaction turnover time for enzymes of central metabolism (*85*).

1.5.2. Biomolecular condensates concentrate enzymes and affect their kinetics

The potential of enzyme complexes to influence metabolism has led to various attempts at the engineering of synthetic enzyme clusters. Most attempts have relied either on using scaffold molecules (protein, lipid, or nucleic acids) to which enzymes are anchored, or the encapsulation of enzymes in bacterially or virally-derived protein microcompartments (*82*). In plants, the metabolon responsible for dhurrin biosynthesis (composed of a UDP-glucosyltransferase and a P450 oxidoreductase) was successfully relocated to the thylakoid membrane by using a TatB/TatC self-assembling scaffold (*86*). This resulted in a 5-fold increase in dhurrin and a decrease in intermediates and side-products in *N. benthamiana*. Cyanobacterial carboxysomes have been the focus of designing protein microcompartments in plants, due to their potential for carbon concentration (*82*), but no increase in Rubisco activity has yet been achieved (*87*). Self-assembling viral or bacterial proteins have been successfully used to encapsulate proteins in eukaryotic systems (*88, 89*).

While these protein-scaffold-based systems have potential, they are complex to engineer, especially as the number of enzymes in a metabolic pathway increases. In plants, the large array of active endogenous proteases may also prove problematic for this approach. The emergence of biomolecular condensates as a widespread organising and regulatory principle in cell biology provides a potentially simpler and more effective solution.

Initially the formation of biomolecular condensates was attributed solely to a phenomenon known as liquid-liquid phase separation (LLPS) (*90*). LLPS is a process whereby a solute demixes into two stable liquid phases, one solute-rich and one solute-poor, because of transient and low-affinity interactions. This occurs as a result of supersaturation, such that above a certain critical concentration (the saturation concentration, c_{sat}) it becomes thermodynamically favourable to partition into two phases; beyond this concentration, an increase in solute doesn't affect the concentration in either phase, but affects the volume occupied by each phase (*91*). However, more recently it has become clear that not all condensates identified in living systems are liquids, but that they can range from gel-like to solid and can even change over time, and may be more accurately described as viscoelastic fluids (*92*). Recent commentaries have noted an assumption in the field that newly discovered condensates are formed through LLPS without the required due diligence needed to establish whether phase-separation is actually occurring (partially due to the difficulty of demonstrating this *in vivo*) (*92, 93*). In addition, much of the work showing these condensates to form by demixing alone has been done *in vitro* or at concentrations well beyond physiological relevance (*91*).

The term ‘liquid-liquid phase-separated compartments’ has hence been superseded by ‘biomolecular condensates’, in recognition of the more complex biophysics underpinning their formation. It has become understood that many more processes beyond simple demixing can cause these condensates to form. For instance, it has been suggested that biomolecular condensates form through a phenomenon termed phase separation coupled to percolation (PSCP), in which biopolymers able to form condensates possess a so-called ‘sticker-and-spacer’ architecture. According to this model stickers mediate specific physical crosslinks and the regions between them, the spacers, will influence the solubility of the polymer and determine the material properties of the condensate (*94*). This model suggests that specific chemical interactions between stickers, and modulated by spacers, are the drivers for the formation of biomolecular condensates, as opposed to the spontaneous and thermodynamically driven processes responsible for LLPS (*93*). Importantly, this model allows for the multiple physical states that have been observed in these condensates including liquids, gels, and solids. Processes such as gelation and aggregation, which have been suggested to be responsible for condensate formation (*95, 96*), can be explained using the same framework as phase-separation. Indeed, several recent computational papers have demonstrated that the difference between phase-separation and aggregation behaviour of a biopolymer is determined by sequence composition and patterning, i.e. by the exact organization of stickers and spacers (*97, 98*).

The molecules which come together to form the condensate are termed the ‘scaffold’; biomolecules such as enzymes or substrates can then be transiently recruited into the

condensate without affecting the condensate itself, and these are referred to as 'clients' (*99*). Recruitment to condensates can thereby concentrate enzymes and their substrates, and can control flux through pathways by selective substrate recruitment (*92*). The stress granules (SGs) are one of the best studied examples of the formation dynamics of biomolecular condensates. SGs form in response to stress and through three steps starting with the formation of a dense core via LLPS (nucleation), followed by the recruitment of clients, and ending with the accumulation of proteins into a less dense peripheral shell. This less dense phase still forms part of the condensate, resulting in a biphasic system (*100*). It is not known to what extent this sequence of events applies to the formation of other condensates, but many examples of condensates with complex phase architectures exist in the literature. This complex organization can take different forms; in the condensate formed by the AUXIN RESPONSE FACTOR (ARF) in *Arabidopsis* root cells, for example, the organization is the inverse of SGs and the stable and dense layer is on the exterior (*101*).

It has long been speculated that concentrating enzymes and substrates together in condensates would increase enzyme efficiency via mass action, but only recently has this been experimentally demonstrated and a mechanism for this enhancement proposed. A seminal paper by Peeples and Rosen used a rapamycin-inducible scaffold-client system to recruit an E2 enzyme and substrates of the SUMOylation cascade to a condensate. They could thereby determine SUMOylation rates as a function of enzyme and substrate concentration in both dense and dilute phases, and calculate kinetic parameters from this data (*78*). In this way a two-part effect of the recruitment of enzymes to condensates on enzyme kinetics was uncovered. Firstly,

they showed that the dense phase accounted for 30 % of the total E2 activity whilst only accounting for 1.1 % total volume, an effect predominantly attributed to the increased enzyme and substrate concentration and thereby higher turnover inside the condensate (i.e. mass action). However, turnover rates when enzyme and substrate concentrations were equivalent to those in the dense phase but LLPS was not induced were 39 % lower; the authors attributed this to the organisational effect of the scaffold, by being able to orient the enzyme and substrates in a favourable way (*102*).

1.5.3. The engineering of biomolecular condensates

It has hence become clear that biomolecular condensates can selectively concentrate enzymes and substrates, and favourably influence enzyme kinetics. It is also known that bringing multiple copies of enzymes forming a pathway together can favourably influence reaction rate through metabolite channelling. The design of synthetic condensates to which single or multiple enzymes are targeted has therefore become an emerging research topic in metabolic engineering.

Many proteins having been shown to phase-separate are intrinsically disordered or have intrinsically disordered regions (IDRs). IDRs usually possess multiple repeated domains, and if these contain residues able to form reversible interactions then these may act as the 'stickers' for LLPS. In IDRs, these interactions are usually charge-charge, dipole-dipole, charge-π, π-π or hydrogen bonds, meaning that tyrosine, arginine or phenylalanine residues are often involved in the LLPS of IDRs (*99*). Glycine and polar amino acid residues do not have strong interaction patterns and thereby act as the spacers (*97*). The LLPS behaviour of certain IDRs has been extensively

investigated *in vitro*; for example, the RGG IDR from the *Caenorhabditis elegans* LAF-1 protein has been shown to stably phase-separate *in vitro* (*103*) and has therefore been used to generate several synthetic RGG-based biomolecular condensates *(104-106)*.

Indeed, the modularity of IDRs for generating biomolecular condensates have led to their exploitation in biotechnology. As mentioned, biomolecular condensates can sequester or exclude biomolecules from the surrounding cytosol thereby resulting in a cellular microenvironment. Synthetic condensates built from IDRs have therefore been proposed as a way of compartmentalising metabolism, for instance by controlling a metabolic branch point. For instance, the formation of a condensate has been combined with optogenetic control to create the optoDroplet system that uses light to trigger compartment formation (*6*). This system relies on the Cry2 domain from *A. thaliana* which oligomerizes upon blue light stimulation. The utility of the system for metabolic engineering was demonstrated by using it to inducibly assemble enzymes of the deoxyviolacein pathway into a compartment using light. These enzymes were fused to the N-terminal IDR from the protein FUS, as well as Cry2, and were shown to form biomolecular condensates in yeast upon blue light stimulation. Localisation of the enzymes to the optoDroplet was shown to yield a 6-fold increase in desired product, by offering control over a metabolic branch point.

As explained recruitment of enzymes to condensates can favourably affect their kinetics. As well as suggesting a reason why biomolecular condensates are so widespread *in vivo*, this also offers intriguing possibilities in biotechnology. Dzuricky and co-workers scanned the sequence space of native intrinsically disordered peptides (IDPs) to design artificial IDPs (A-IDPs) (*107*). A-IDP condensates were

created in both eukaryotic (human embryonic kidney 293 cells) and prokaryotic (*E. coli*) systems and the α peptide of β-galactosidase was fused to A-IDP and could complement the mutated enzyme LacZΔM15 to create a functional β-galactosidase enzyme inside the condensate. The results showed a statistically significant increase in the catalytic constant or turnover efficiency (k_{cat}) of the enzyme with increasing A-IDP molecular weight. This system offers the advantage of presenting a complete library of IDPs with a range of well-characterised biophysical properties and could be used to rapidly scan IDP targets for ability to cause phase-separation or to sequester an enzyme of interest. Similarly, a library of phase-separation-promoting peptide building blocks was developed to allow for the tuning of phase-separation behaviour, dynamics, and client recruitment efficiency. The increased knowledge regarding interaction between scaffold structure and phase separation will help with the design of compartments tailored to specific applications (*108*).

The scaffold-client system known to operate *in vivo* can be repurposed for biotechnology. In many cases, to generate synthetic condensates and recruit enzymes, an IDP is fused to the enzyme to be localized to the compartment. However, fusing large disordered domains to proteins could alter their functionality, and concerns have been raised about the generalizability of this approach (*109*). Good and colleagues have developed a platform combining high client recruitment to the condensate and a strategy of localizing the clients without disrupting their native function. To do this they designed two constructs; the first consisted of two or three copies of the RGG domain fused on the N-terminus to a member of a coiled-coil pair. The second construct consisted of the client enzyme fused on its C-terminus to the

other member of the coiled-coil pair. Strikingly, they showed the coiled-coil scaffold was able to recruit over 90 % of the client tagged with the cognate interacting motif in *S. cerevisiae* (*109*).

Finally, as well as the determination of enzyme kinetics inside biomolecular condensates, the paper by Peeples and Rosen already described uncovered selective effects on recruited enzymes, with implications for the design of synthetic condensates. For example, a greater rate increase was observed for enzymes with a high K_m, as enzymes with a low K_m are likely already operating close to maximum efficiency in dilute conditions. In addition, enzymes experiencing substrate inhibition might be negatively affected by recruitment to condensates due to higher substrate concentrations (*78*).

1.6. Computational modelling can provide insights into metabolic networks

Given the complexity of metabolism, we cannot rely on intuitively deduced solutions to metabolic engineering problems (*5*). As an example, consider one facet of metabolic engineering already mentioned, namely substrate availability. In bacterial PHB production systems, a large proportion of ancillary engineering interventions (i.e. interventions on top of the PHB biosynthetic pathway) are concerned with acetyl-CoA provision (*110*). In the cytosol, ATP-citrate lyase (ACL) catalyses the ATP and CoA-dependent cleavage of citrate to form acetyl-CoA and oxaloacetate. In *A. thaliana,* ACL is a heterooctamer formed of the two subunits ACLA and ACLB, encoded by three and two genes, respectively, and could be an attractive engineering target to increase

provision of cytosolic acetyl-CoA (*111*).

However, using ACL to drive production of acetyl-CoA for cytosolic PHB synthesis is likely to have some far-reaching metabolic consequences. For example, the cytosolic elongation of C18 fatty acids for the formation of longer chain fatty acids used in storage triacylglycerols is dependent on the provision of acetyl-CoA in the cytosol by ACL (*112*). A metabolic flux analysis of developing *Brassica napus* embryos showed that essentially all the citrate formed in the mitochondrion is exported and is used for cytosolic acetyl-CoA production by ACL. This analysis also revealed some unconventional mitochondrial metabolism, such as the absence of flux around the TCA cycle (*113*). Cytosolic acetyl-CoA generated by ACL has also been shown to be essential in cell wall polysaccharide (*114*) and protein (*115*) acetylation. The overexpression of ACL in cytosolic PHB-producing *A. thaliana* lines was able to rescue phenotypes associated with PHB synthesis, but also significantly impacted wax and cutin biosynthesis, revealing the complexity of ACL regulation (*61*).

Hence metabolism does not operate as the sum of clearly defined metabolic pathways; instead, these pathways do not always follow a canonical flux pattern and form part of a larger interconnected network. Computational models of the whole metabolic network have become an important tool to determine and rapidly test the effects of metabolic engineering strategies. Metabolic network models can be built from genome sequence data and relies on an annotated genome from which to obtain the EC numbers of genes shown to code for enzymes. These EC numbers can be used to search databases such as KEGG (*116*) or BioCyc (*117*) to obtain reactions catalysed by these enzymes. The quality of a metabolic model and the feasibility of building it is

therefore limited by the quality and availability of annotated genome sequences.

The modelling of metabolic networks usually takes one of two approaches. In the first, kinetic models, the system can be numerically simulated for a given starting concentration of metabolites if the reaction stoichiometry, rate equation, and kinetic parameters of each enzyme in the system are provided. Kinetic models thereby provide a quantitative description of metabolic phenotypes by linking reaction rates, enzyme levels, and metabolite concentrations, and are therefore predictive (*9, 118*). As such more complicated metabolic interventions than a simple knockout or upregulation can be modelled, such as changes in enzyme levels, regulatory effects, or metabolite pools. For example, kinetic models can be used to identify allosteric regulation mechanisms under different metabolic conditions. In *Clostridium thermocellum*, a kinetic model was used to identify metabolic changes associated with nitrogen limitation and leading to increased lactate, pyruvate, and amino acid yields, as well as effects of ethanol stress on sugar phosphate concentrations (*119*). In addition, kinetic models can distinguish between isozymes, and was used to identify a hexokinase isozyme controlling xylose uptake as a knockout target in *S. cerevisiae*; this target was subsequently experimentally shown to increase xylose uptake rates (*120*).

Despite the power of kinetic models, their widespread application has been hampered by a lack of data. The quality of the model is ultimately dependent on how much kinetic data is available to parameterise it as well as the accuracy of this data; metabolic systems in which the kinetics of all enzymes have been characterised are very rare, and in most cases information will be drawn from different species or a combination of *in vivo* and *in vitro* enzyme characterisations. In plants, the dearth of information has

meant that kinetic models have been confined to subsets of metabolism such as the Calvin-Benson-Bassham cycle (*121*). In addition discrepancies between *in vivo* and *in vitro* parameter values are frequent and can be significant; the k_{cat} values for the malate dehydrogenase from *E. coli* differ by 100-fold (*122*).

Given that *in vivo* parameters are generally considered a more accurate reflection of the activity of the enzyme (*9*), many techniques have been explored for the derivation of *in vivo* kinetic parameters from different types of data. For example, *in vivo* labelling experiments can be used to derive enzyme kinetics. Briefly, an isotopically labelled precursor, usually using ^{13}C, is supplied and a time series is obtained. Enzyme kinetics which would lead to the observed labelling kinetics can be determined using data-fitting techniques. This kinetic parameterisation often involves the integration of a large system of non-linear ordinary differential equations, and the non-linearity of the system poses problems for the optimisation problem such that long integration times or the inability to converge on a solution are common issues (*123*). Recently, a machine learning tool allowing for the prediction of enzyme kinetic parameters from protein sequences and substrate structures, and the effect of environmental factors, was published (*124*). Resources based on machine learning such as this one could play an important role in building complete and accurate kinetic models of metabolism.

The second approach to metabolic modelling is known as stoichiometric modelling. Stoichiometric modelling relies solely on knowledge about the structure of the metabolic network and the stoichiometry of the metabolic reactions defined in the model, which can be extracted from genome sequence data. In order to reduce the

possible flux solution space, constraints such as reaction reversibility or stoichiometry can be imposed, and a steady state assumption applied. If an objective function is imposed, i.e. a reaction which should be maximized, linear programming can be used to find a solution which satisfies this objective within the constrained solution space (*9*). This solution is a set of metabolic flux distributions which are optimal with respect to the objective; these flux distributions are seldom unique, however, a problem which can be tackled by applying a secondary optimization criterion such as minimization of overall fluxes (parsimonious flux balance analysis [pFBA]) (*125*). Because of the steady state constraint often applied in this type of modelling it has become known as 'flux balance analysis' (FBA). Using this approach large networks of several thousand enzymes can be modelled and can still, despite their relative simplicity, be used predictively in certain instances. For example, if one assumes that knocking out a gene results in the rearrangement of metabolism so as to minimize change, the effect of gene mutation on metabolic fluxes can be accurately predicted (*126*). Stoichiometric models are also very powerful tools to fully explore the endogenous network, such as alternative routes and pathways, which can be important in the design of metabolic engineering strategies (*5*). Tools have been specifically developed for the exploration of alternative pathways in stoichiometric models, such as flux sampling (*127*).

Although the simplicity of stoichiometric models facilitates their construction as well as reduces computational load, it does also have some drawbacks. For instance, stoichiometric models do not incorporate knowledge regarding metabolic network regulation. In addition, and specifically with regards to plants, the accuracy of stoichiometric models is limited by knowledge regarding metabolic

compartmentation. Knowledge regarding the subcellular localization of enzymes as well as integral membrane transporters remains incomplete, although some progress has been made using organelle-specific proteomics (*128*), and the characterization of transporters (*129*). Several approaches have been devised to integrate additional information into stoichiometric models in order to increase their predictive power. For instance, some enzyme kinetic information can be incorporated into stoichiometric models, and has been shown to substantially improve predictive accuracy in a small scale hybrid model if applied to enzymes at network branch points (*130*). Introducing even a small number of kinetic expressions, however, converts the linear optimization problem to a non-linear one, leading to an increase in computation time. The same is true of efforts to introduce the effects of environmental parameters such as light or humidity (*131*). In addition, stoichiometric models have been integrated with 'omics datasets and combined with machine learning to improve model predictability and thereby improve the design of metabolic engineering strategies (*125*).

As described FBA usually involves solving a linear optimization problem at steady state given an objective function. When modelling endogenous metabolism, the objective is usually biomass accumulation. In an engineered organism, however, the production of biomass often needs to be balanced with target molecule production. The OptKnock framework solves for optimal flux distribution given the simultaneous optimization of biomass accumulation and secretion of a target metabolite (*132*). When experimentally implemented, three gene deletion strategies devised using OptKnock resulted in significantly increased lactic acid production in *E. coli* coupled to growth rate, i.e. an increase in lactic acid production also resulted in increased

biomass (*133*). However, in engineered cells it may not be accurate to assume that metabolism is configured so as to optimize growth given the often suboptimal growth and metabolite levels of engineered organisms (*134*). To address this a different objective function can be used, such as the minimization of metabolic adjustment (MOMA), which tries to minimize the difference between the wild-type and engineered flux profile thereby returning a flux distribution closest to the optimal flux distribution in the wild-type (*135*). This approach has been used to improve the yield of lycopene from *E. coli* (*134*).

1.7. Thesis rationale and aims

The combination of complexity, unpredictability, and tight regulation of central metabolism, as well as slow iteration through metabolic changes, means that establishing a plant-based production system with high product titre is exceedingly difficult. What is more, the economic viability of plant-based production systems is highly yield-dependent (*136*). Even in simpler and better understood organisms such as bacteria, developing new cell factories with the economic requirements needed for industrial scale production typically requires 6-8 years and over 50 million USD (*8*). As such there is an urgent need in plant metabolic engineering for ways to overcome these challenges and in doing so increase product yield of target molecules. This problem remains a big obstacle to the establishment of plants as viable bioproduction platforms in the green economy.

Computational models have proven useful in metabolic engineering by revealing unexpected and unintuitive effects on the larger metabolic network caused by an

engineering intervention. Kinetic models of metabolism are generally more predictive than stoichiometric models but are limited by data availability as well as long computation times. Stoichiometric models can be constructed for any species for which there is a fully annotated genome, but for which enzyme kinetic parameters might not have been determined, and can be used to investigate changes to network structure upon metabolic perturbation such as the introduction of heterologous genes.

In this project a stoichiometric model was constructed for the liverwort *M. polymorpha* and used to investigate the metabolic engineering of this organism. *M. polymorpha* has been used for nearly 200 years to study the effect of environmental factors on morphology and physiology (*137*). Although the development of genetic and genomic tools in the 20th century favoured other model species such as *A. thaliana*, more recently developed tools for transformation as well as genomic resources have revived interest in *M. polymorpha* as a modern model with benefits including short life cycle, ease of propagation, efficient transformation, haploidy, and a genome of only ~280 Mb (*137, 138*). Importantly, primary metabolism has been shown to be essentially conserved in the green lineage (*139*), meaning engineering interventions tested in *M. polymorpha* should be translatable to higher plants.

One of the challenges in plant metabolic engineering is that the production of a molecule of interest often competes with endogenous metabolism, leading to unintended and negative effects on phenotype. Inducibility can be used to minimize effect of production on metabolism by controlling when and how expression occurs. In this project I will test an optogenetic induction system known as PULSE in *M.*

polymorpha. Optogenetic induction has potential in being less leaky and having less crosstalk with natural plant processes than chemical inducers.

In addition, compartmentation is used throughout biology to control and accelerate metabolic flux. In the last 10 years a large body of research regarding biomolecular condensates has emerged, and details regarding the control over metabolic processes and effects on enzyme kinetics of these structures have been characterised. Using this knowledge synthetic biomolecular condensates have been generated in a range of organisms. In this thesis synthetic biomolecular condensates will be implemented in plants for the first time and their usage as a tool in metabolic engineering investigated.

In all cases the exploration of metabolic engineering challenges in *M. polymorpha* will be illustrated using PHB as a case study. The production of PHB in plants has been well-studied but has suffered many of the drawbacks commonly seen in plant metabolic engineering, resulting in poor yields. For instance, negative growth phenotypes associated with high yields of PHB have been associated with competition between PHB accumulation and endogenous metabolism for acetyl-CoA. The PULSE system will be linked to the production of PHB in order to investigate the ability of optogenetics to alleviate competition effects between PHB and biomass accumulation. In addition, particular aspects of the PHB biosynthesis pathway lends itself to metabolic branch point control which could be achieved using compartmentation into biomolecular condensates. The stoichiometry of β-ketothiolase (involving two acetyl-CoA molecules, Figure 1.1A) is likely to make it more sensitive to decreased acetyl-CoA levels (as long as the concentration is below the enzyme saturation level), an argument put forward to explain why citrate synthase

outcompetes β-ketothiolase under exponential growth of modelled bacteria (*62*). Controlling the metabolic branch point at acetyl-CoA could be used to increase flux through PHB synthesis, although a dedicated source of acetyl-CoA may be needed to prevent even more severe negative phenotypes. Computational modelling of PHB production in *M. polymorpha* will reveal this and other key changes to central metabolism required for maximal PHB production.

To summarise, the key aim of this research project is to establish whether *M. polymorpha* is a useful testbed for testing metabolic engineering strategies, using the production of PHB as a case study. This will be achieved by the following specific research objectives:

- To construct a computational model of *M. polymorpha* metabolism and to analyse the model to inform the metabolic engineering strategy for PHB production (Chapter 3)
- To implement inducible PHB biosynthesis in *M. polymorpha* using an optogenetic system, and to compare this to constitutive PHB production (Chapters 4 and 5)
- To explore ancillary metabolic engineering strategies identified using modelling to optimise PHB production (Chapter 5)
- To investigate the targeting of PHB pathway enzymes to biomolecular condensates for enhanced PHB production (Chapter 6)

2. Materials and methods

2.1. Computational modelling

2.1.1. Model construction

A stoichiometric model representing the core metabolism of *M. polymorpha* was built based on an *Arabidopsis* model as explained in previous work (*140*). Briefly, the *Arabidopsis* model used as a starting point (PlantCoreMetabolism_v1_2) is a fully compartmentalised, mass-balanced, and charge-balanced model of the conserved reactions of plant primary metabolism (*141*). To ensure the model accurately represented the core metabolism of *Marchantia*, a homology search of *Arabidopsis* genes in the *Marchantia* genome was carried out using OrthoFinder (*142*) and led to the removal of five reactions (MetaCyc identifiers RXN-1121, RXN-10773, MYO-INOSITOL-OXYGENASE-RXN, CONIFERIN-BETA-GLUCOSIDASE-RXN and 2.4.1.111-RXN). Fifty seven reactions and 56 metabolites were added to account for long-chain fatty acids present in *Marchantia* but not in *Arabidopsis*.

Predicting rates of biomass accumulation requires a so-called 'biomass equation' which combines individual biomass components in the correct relative stoichiometries based on experimentally determined values for cellular biomass composition. To generate a biomass equation for *M. polymorpha* thallus tissue, biomass composition information was compiled using an in-depth literature search (*143-149*). In the case of total protein and soluble metabolite contents and composition, information was not available for *Marchantia* and values from

Arabidopsis were used. Hemicellulose and pectin polymers were represented by xylan and homogalacturonan, respectively. The biomass composition and component stoichiometry in the biomass equation as well as the reactions added to the model can be found at https://github.com/AnyaLindstromBattle/Thesis.

2.1.2. Estimation of non-growth associated maintenance (NGAM) costs

The energetic costs for NGAM were modelled by the addition of constraints to modulate flux through ATP hydrolysis and NADPH oxidase reactions, as previously described (*140*). Specifically, the flux through ATP hydrolysis and NADPH oxidase was constrained in a 3:1 ratio, shown to lead to the flux distributions using FBA most closely resembling those observed *in vivo* (*150*). The value for NGAM in *M. polymorpha* was determined based on data for a mature photomixotrophic culture of *M. polymorpha* cells (*151*). The calculated value was used to constrain flux through the ATP hydrolysis reaction. NGAM costs were assumed to be equal between night and day.

2.1.3. Setting up a diel model of *M. polymorpha*

The model was set up to be able to optimise flux over both day and night in a single optimisation using the so-called diel FBA framework (*150*). Two replicate copies of the model were created, one representing day- and the other night-time metabolism, and identifiers added to metabolites and reactions to indicate their assignment to the day or to the night phase. Reactions to allow diel storage (i.e. storage at night for use during the day or vice versa) of starch, amino acids, and soluble metabolites were added. In addition, pseudometabolites (nitrate uptake, Rubisco carboxylation:oxygenation, and

biomass output) were added to constrain day- and night-time reactions to specific flux ratios; these values were assumed to be identical to *Arabidopsis* given a lack of *Marchantia*-specific data and are shown in Table S2.1.

2.1.4. Further model constraints

The model of *M. polymorpha* previously described *(140)* was further modified in this project to generate a more accurate model of *M. polymorpha* metabolism. Constraints were added on the diel transfer of malate, citrate, and fumarate in the vacuole, based on experimental results obtained for *M. polymorpha* of the diurnal changes in these carboxylic acids (Corinna Hartinger, University of Oxford, unpublished data, see https://github.com/AnyaLindstromBattle/Thesis). To model the carbon sink of the rhizoids, sucrose transport out of the thallus model was allowed, and assumed to occur in a 3:1 ratio between day and night (same as the ratio of day:night phloem export in *Arabidopsis* (54)). Flux to the rhizoids was constrained such that it carried 4.3 % of the carbon assimilated by the model, based on data from *Polytrichum commune* (*152*).

Finally, the model was modified to allow for the accumulation of PHB in both the cytosol and the plastid by adding the reactions outlined in Table 2.1. All constraints are summarized in Table S2.1.

Table 3.1: Reactions added to the baseline *M. polymorpha* model to allow for accumulation of PHB in the cytosol and in the plastid. Reactions to allow PHB to drain from the system were also added but not shown. '_c' indicates a cytosolic reaction/metabolite whereas '_p' indicates a plastidial reaction/metabolite. Reactions were obtained from MetaCyc (117). CPD_650: 3-hydroxybutyryl-CoA.

Reaction name	Reaction	Reasoning
Phb_A_c	2 ACETYL_COA_c <=> ACETOACETYL_COA_c + CO_A_c	The production of PHB occurs via the concerted action of β-ketothiolase (PhbA), acetoacetyl-CoA reductase (PhbB), and PHB synthase (PhbC) (Figure 1.1). Acetoacetyl-CoA reductase exists as two versions; the *A. thaliana* version makes use of NAD, but the R. *eutropha* version (used for most PHB engineering attempts) uses $NADP^+$ (*153*).
Phb_A_p	2 ACETYL_COA_p <=> ACETOACETYL_COA_p + CO_A_p	
Phb_B_c	ACETOACETYL_COA_c + NADPH_c + PROTON_c <=> CPD_650_c + NADP_c	
Phb_B_p	ACETOACETYL_COA_p + NADPH_p + PROTON_p <=> CPD_650_p + NADP_p	
Phb_C_c	CPD_650_c <=> CO_A_c + POLY_HYDROXY_BUTYRATE_c	
Phb_C_p	CPD_650_p <=> CO_A_p + POLY_HYDROXY_BUTYRATE_p	
NphT7_c	ACETYL_COA_c + MALONYL_COA_c + PROTON_c --> ACETOACETYL_COA_c + CARBON_DIOXIDE_c + CO_A_c	The engineering of PHB production in plants has been achieved with the first step catalysed either by acetoacetyl-CoA synthase (*NphT7*) or β-ketothiolase (*46*).
NphT7_p	ACETYL_COA_p + MALONYL_COA_p + PROTON_p --> ACETOACETYL_COA_p + CARBON_DIOXIDE_p + CO_A_p	
ACETYL_COA_CARBOXYLTRANSFER_RXN_c	ACETYL_COA_c + 0.65 ATP_c + HCO3_c + 0.35 aATP_c --> 0.5 ADP_c + MALONYL_COA_c + 0.55 PROTON_c + 0.7 Pi_c + 0.5 aADP_c + 0.3 aPi_c	No cytosolic malonyl-CoA exists in the baseline model, so a way of making it has to be added. In *A. thaliana*, malonyl-CoA is made almost exclusively from acetyl-CoA by acetyl-CoA carboxylase (ACC) (*154-156*). Whether a cytosolic ACC exists in *M. polymorpha* is unclear.

2.1.5. FBA and model analysis

The workflow for FBA has been previously described (*157*). FBA was performed using the CobraPy package (version 0.29.0) (*158*) installed on Python (version 3.8.18, Python Software Foundation, https://www.python.org/). The pFBA function (*159*) available in CobraPy was used to perform linear optimization using the GLPK solver (http://www.gnu.org/software/glpk). To compare flux distributions between models set up differently, flux variability analysis (FVA) (*160*) was used to determine the maximum and minimum possible value of the flux through each reaction and fluxes considered different if the ranges did not overlap.

2.1.6. Flux sampling

The OPTGP flux sampling algorithm (*161*) available in CobraPy was used to generate sets of 5,000 samples from different model setups. Sampling was run using different thinning constants; sample chains with thinning applied will store every k^{th} sample, where k is the thinning constant (*162*). This can help avoid the effects of autocorrelation (where samples close together in a chain can be similar) and ensure convergence. Four different thinning constants (k = 1,000, 10,000, 100,000, 1,000,000) were used and three convergence diagnostics were computed for sample chains using the CODA package in R (*163*) and only converged samples were used for further analysis. Convergence was confirmed by inspecting trace and autocorrelation plots of individual reactions.

Flux distributions for each reaction across the solution space of two models were compared to identify alternative pathways. Reactions where fluxes were always < 1^{-6}

μmol m^{-2} s^{-1} were excluded from the analysis. Similarly, reactions with distributions showing unrealistically high absolute fluxes (where flux distributions reached the maximum or minimum flux value allowed in the model [1,000 and -1,000 μmol m^{-2} s^{-1}, respectively]) were ignored because they were found to always be part of futile cycles. This arises because flux sampling does not include the secondary objective of minimization of fluxes used in pFBA, nor does it constrain carbon assimilation to an experimentally derived value as explained in Table S2.1. As such some flux distributions contain solutions with very high flux values which are purely the result of higher carbon and energy availability; the futile cycles are used by the model to dissipate the excess energy.

Reactions with (a) a significant ($P < 0.05$, Kruskal-Wallis (*164*)) difference between the two flux distributions, and (b) a difference between the point of highest density in the distribution and the pFBA result of flux through that reaction were taken forward for further analysis using a constraint scan and a correlation analysis. Briefly, each reaction satisfying the two conditions mentioned was constrained to five different flux values evenly spaced between the first and the third quantile of its sampling distribution. For each value pFBA was run and Pearson's correlation coefficient (r) calculated for flux through the reaction across the five points tested against flux through each of the other reactions in the model. Reactions with non-zero r and slope were considered as possible candidates in a pathway including the reaction which was varied.

2.1.7. Data availability

The stoichiometric model of *M. polymorpha* in SBML format as well as Python functions needed to run the model and associated information can be found at https://github.com/AnyaLindstromBattle/Thesis.

2.2. Data analysis and statistics

Statistical analysis and data visualisation were performed using the Python programming language (version 3.8.18, Python Software Foundation, https://www.python.org/) in a JupyterNotebook (*165*). Statistical differences between groups were determined using different statistical tests depending on the application. To determine differences between distributions a Kruskal-Wallis (*164*) test was used; for differences between multiple groups an ANOVA was followed by a *post-hoc* Tukey's test (*166*); for differences between two groups a Mann Whitney U (*167*) test was used. All tests were implemented using the SciPy package (*168*). Figures were prepared using the seaborn (*169*) and matplotlib (*170*) packages. For principal component analysis (PCA) I used the scikit-learn package (*171*).

The specific statistical test used in each analysis and n values are described in figure legends, and error bars represent standard error of the mean (SEM). Significance was reported as '***' for a P-value < 0.001, '**' for a P-value < 0.01, or '*' for a P-value < 0.05.

2.3. Plant materials and growth

N. benthamiana was grown in a greenhouse with a 16:8 h light:dark cycle at 25 °C. Natural light was supplemented with ATTIS-7 LED grow lights (Plessey, London, UK) to

achieve a light intensity of up to 200 µmol photons m^{-2} s^{-1}. Plants were grown in 11-cm square pots on soil composed of a 3:1 ratio of modular seed growing compost and pro fine vermiculite (both Sinclair Pro, Cheshire, UK) and were regularly watered. 0.4 g L^{-1} Exeptor (ICL, Ipswich, UK) was used as an insecticide.

Male and female accessions of *M. polymorpha* (Takaragaike-1 [Tak1] and Takaragaike-2 [Tak2], respectively (*35*)) were mostly grown aseptically on 12-cm square petri plates on either ½ Gamborg B5 media (0.158 % [w/v] Gamborg B5 + vitamins [Duchefa, Haarlem, Netherlands], 2.34 mM MES monohydrate, 29.2 mM sucrose, 1.4 % [w/v] plant agar, pH 5.6), or ½ MS media (0.22 % [w/v] MS salts + vitamins [Duchefa, Haarlem, Netherlands], 29.2 mM sucrose, 0.8 % [w/v] plant agar, pH 5.7) in the case of thallus transformation. Plants were grown in a growth room with a 16:8 h light:dark cycle at 22 °C. In the case of the PULSE transformants and when reproductive growth was required for crossing, plants were grown in a growth room with a 16:8 h light:dark cycle at 18/22 °C day/night temperature with the addition of 10 µmol photons m^{-2} s^{-1} far-red light (see growth room spectra in Figure S2.1). Prior to inducing reproductive growth, 3–4-week-old plants were transferred to soil (2:1 ratio of John Innes Number 2 [Bulrush, Magherafelt, UK] and fine vermiculite [Sinclair Pro, Cheshire, UK], autoclaved) in autoclaved 1 L microboxes ($SacO_2$, Deinze, Belgium).

To generate *M. polymorpha* spores for transformation, plants were grown in white light supplemented with far-red light as described for around 5 weeks until reproductive structures emerged. A Pasteur pipette was used to place a drop of water on the antheridiophore for 5 min after which the water was resuspended several times to release sperm before being placed on an archegoniophore ready for fertilisation.

Sporangia emerged 4-5 weeks after crossing and were dried for 10 d in a 1.5 mL tube with a filter on top in an air-tight box containing silica gel. Typically, five or more sporangia were collected per tube; the contents of one tube were then split between two transformations. After drying, spores were transferred to -80 °C for long-term storage.

Light spectra and intensities were measured using an AvaSpec-ULS2048CL-EVO spectrometer (Avantes, Apeldoorn, Netherlands). Blue light was defined as 420 – 490 nm, red light as 620 – 680 nm, and far-red light as 700 – 750 nm.

2.4. Molecular cloning and genetic construct fabrication

2.4.1. Genetic construct design

All gene parts used in genetic construct assembly can be found in Table 2.2. Most gene sequences were chemically synthesized by Twist Bioscience (San Francisco, USA) and supplied in the Gateway entry vector pTWIST-Entr. The exceptions to this were the *Malus domestica* citramalate synthase (*Md*CMS) which was obtained from a plasmid designed by András Sándor (University of Oxford, UK) (*172*), the chimeric homomeric ACL *At*ALB (from Yves Poirier, Université de Lausanne, Switzerland) (*61*), the PULSE optogenetic system (from Uriel Urquiza and Matias Zurbriggen, Heinrich Heine University, Düsseldorf, Germany) (*73*) and the green fluorescent monomeric protein mClover3 (*173*) as well as the hygromycin resistance cassette (both from Francesco Licausi, University of Oxford, UK). Gateway expression vectors p*Mp*GWB103 and p*Mp*GWB303 (Addgene plasmids numbers 68557 and 68631, respectively) (*42*) were a gift from Takayuki Kohchi from Kyoto University, Japan. The Gateway expression vector

pK7WG2 (*174*) was a gift from Francesco Licausi. Plasmids were designed and assembled *in silico* using SnapGene (version 7.1.0, GSL Biotech, San Diego, USA). Synthesized sequences were codon optimized for the host species using either the in-house codon optimization software provided during sample submission to TWIST (for *N. benthamiana*), or by using OPTIMIZER (*175*) and a codon usage table obtained from the Codon Usage Database (http://www.kazusa.or.jp/codon/) (for *M. polymorpha*). The 'guided random' approach had to be used with OPTIMIZER to avoid repetitive sequences incompatible with gene synthesis (*175*).

Table 2.2: Gene parts used for the design of genetic constructs.

Name	Description	Sequence source or protein sequence
PULSE	PULSE is an optogenetic system that is inactive under white light and activated with monochromatic red light (*73*). Carries resistance to kanamycin. For more information see section 4.3.1.	Part ID JPUB_016925 from the Joint BioEnergy Institute ICE repository
HygR	Hygromycin resistance cassette	Addgene plasmid number 71536
PhbA	Coding sequence for β-ketothiolase from *R. eutropha*	Genbank accession number FJ897461.1
PhbB	Coding sequence for acetoacetyl-CoA reductase from *R. eutropha*	Genbank accession number FJ897462.1
PhbC	Coding sequence for PHB synthase from *R. eutropha*	Genbank accession number MH558939.1
P2A	Self-cleaving 2A peptide from porcine teschovirus-1 2A (*176*). 2A peptides are 18-22 amino acid viral oligopeptides which mediate the cleavage of polypeptides during translation in eukaryotic systems, by the ribosome skipping of the formation of a glycyl-prolyl peptide bond at its C-terminus (*177*).	(*178*)
T2A	Self-cleaving 2A peptide from *Thosea asigna* virus 2A (*176*).	(*178*)
InteinF2A	Self-cleaving fusion protein domain composed of a *Ssp* DnaE mini-intein variant engineered for hyper-N-terminal autocleavage which is covalently linked to the F2A peptide (self-cleaving 2A peptide from foot-and-mouth disease virus) (*179*).	(*179*)
GSG linker	A glycine-serine-glycine (GSG) linker fused to the N-terminus of 2A peptides which can improve cleavage efficiency.	(*178*)
ACLY3KQ	The protein sequence of human ATP citrate lyase	UniProt entry

	(ACLY) was reverse translated using the Sequence Manipulation Suite (version 2) from bioinformatics.org (*180*) using a codon usage table for *M. polymorpha* from the Codon Usage Database (http://www.kazusa.or.jp/codon/).	number P53396-2 with the modifications explained in section 5.4.1.
***At*tDT**	The tDT transporter from *A. thaliana* is a probable vacuolar citrate exporter/malate importer (*181*). The protein sequence was reverse translated as described for ACLY3KQ.	UniProt entry number Q8LG88
***At*ALB**	Chimeric ACL protein generated by linking the ACLA1 and ACLB2 subunits from *A. thaliana* ACL with a linker region from *Drosophila melanogaster* ACL.	(*61*)
RGG	IDR from the *C. elegans* protein LAF-1	(*182*)
SYNZIP1 and SYNZIP2	Heterospecific synthetic coiled-coil peptide pair	(*183*)
mCherry	Monomeric red fluorescent protein	Genbank accession number AY678264.1
mClover3	Monomeric green/yellow fluorescent protein	Genbank accession number ATE88096.1
***Md*CMS**	Citramalate synthase enzyme MdCMS_1 from *Malus x domestica*.	(*184*)
FLAG	Short hydrophilic protein tag used to detect proteins in Western blotting	DYKDDDDK
His	Polyhistidine tag enabling purification of the constructs using Ni-affinity purification.	HHHHHHHH or HHHHHHHHHH
Myc	Short protein tag derived from the c-myc gene product and used to detect proteins in Western blotting	EQKLISEEDL
HA	Short protein tag derived from the human influenza hemagglutinin surface glycoprotein and used to detect proteins in Western blotting	YPYDVPDYA

2.4.2. Genetic construct assembly

DNA amplification using polymerase chain reaction (PCR) was carried out using either

Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) or Q5 High-Fidelity DNA Polymerase (New England Biolabs, Hitchin, UK) according to the manufacturer's instructions. Primers used for sequencing and cloning were designed using SnapGene or the NEBuilder Assembly Tool (New England Biolabs, Hitchin, UK) and synthesized by Merck (Darmstadt, Germany). Primers were checked for unique binding sites using SnapGene (version 7.1.0, GSL Biotech, San Diego, USA), for absence of secondary structures and self-dimerisation using the OligoAnalyzer Tool (Integrated DNA Technologies, Coralville, Iowa, USA), and for melting temperature compatibility using the T_m Calculator (New England Biolabs, Hitchin, UK). Primer information can be found in Table S2.2. After PCR fragments were purified using either the QIAquick PCR purification kit or the QIAquick gel extraction kit (both Qiagen, Hilden, Germany). In the case of the former the fragments were also digested using DpnI restriction enzyme to remove template plasmid (Thermo Fisher Scientific, Waltham, USA). PCR fragments were assembled for transformation using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Hitchin, UK). Construct assembly occurred into the pTWIST-Entr Gateway cloning vector, and the Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific, Waltham, USA) used to move the constructs into one of several Gateway expression vectors according to the manufacturer's instructions. Plasmids used in this study are shown in Table S2.3.

2.4.3. Transformation of *E. coli*

A standard heat-shock protocol was used for transformation of *E. coli*. Briefly, approximately 50 ng of DNA was added to an aliquot of α-select Silver Chemically Competent Cells (Meridian Bioscience, Cincinnati, USA) and the cells were incubated

on ice for 30 min. The cells were heat-shocked at 42 °C for 45 sec before incubation on ice for another 10 min. 500 µL Lysogeny Broth (LB) media (10 g L^{-1} tryptone, 10 g L^{-1} NaCl, 5 g L^{-1} yeast extract) was added and cells were incubated with shaking for 60 min at 37 °C or 90 min at 30 °C depending on the presence of repetitive sequences. Cells were plated onto solid LB agar (LB media supplemented with 15 g L^{-1} agar) containing appropriate antibiotics and incubated at 37 or 30 °C for 1-2 d.

Competent *E. coli* cells were prepared in-house using the rubidium chloride method (*185*). Ninety nine mL LB was inoculated with 1 mL overnight culture in a 250 mL flask and grown until OD_{600} reached 0.4-0.55. The culture was chilled on ice for 10-15 min before centrifugation at 2700×g for 10 min at 4 °C. Cells were gently resuspended in 33 mL chilled RF1 buffer (100 mM RbCl, 50 mM $MnCl_2{\cdot}4H_2O$, 30 mM potassium acetate, 10 mM $CaCl_2{\cdot}2H_20$, 15 % [v/v] glycerol) and incubated on ice for 1 h. Cells were pelleted at 580×g for 15 min at 4 °C before resuspension in 8 mL chilled RF2 buffer (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2{\cdot}2H_20$, 15 % [v/v] glycerol) and incubation on ice for 15 min. 100 µL aliquots were dispensed into chilled sterile 1.5 mL tubes before freezing in liquid nitrogen (LiN_2) and storage at -80 °C.

2.4.4. Isolation and analysis of plasmid DNA

For isolating plasmid DNA, individual colonies were grown overnight at 37 or 30 °C with shaking in 5 mL LB media supplemented with the appropriate antibiotics. DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. At least five colonies per construct were screened using diagnostic restriction digestion with an appropriate FastDigest restriction

enzyme (Thermo Fisher Scientific, Waltham, USA) followed by Sanger sequencing (Source Bioscience, Nottingham, UK) and full plasmid sequencing (Plasmidsaurus, Eugene, USA) to select the clone with which to proceed.

2.4.5 Transformation of *A. tumefaciens*

Confirmed plasmids were transformed into *A. tumefaciens* strain GV3101 (pMP90) (*186*) courtesy of Helen Prescott (University of Oxford, UK) using a freeze-thaw protocol. 1,000 ng of plasmid DNA was incubated with an aliquot of competent cells for 5 min on ice. The cells were then flash frozen in LiN_2 for 5 min before thawing at room temperature for 10 min. Two mL LB media was added and the cells incubated with shaking at 28 °C for 4 h. Cells were pelleted at 2700×g for 2 min, resuspended in 450 µL LB and plated onto LB plates containing appropriate antibiotics. Antibiotic resistant *A. tumefaciens* colonies were screened using colony PCR with appropriate primers to bind the construct and PCRBIO Taq Mix Red (PCR Biosystems, London, UK).

A. tumefaciens competent cells were prepared as follows: three 250 mL flasks containing 150 mL LB were inoculated with 1 mL overnight *A. tumefaciens* culture and incubated at 28 °C with shaking until OD_{600} reached 0.8-1.0. Cultures were combined and chilled on ice for 15 min before distributing equally into eight 50 mL pre-cooled tubes. Cells were pelleted by centrifugation at 4,000×g for 20 min at 4 °C, the supernatant discarded, and pellet resuspended in 10 mL ice cold H_2O. All tubes were made up to 50 mL with ice cold H_2O before pelleting at 4,000×g for 20 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 10 mL ice cold H_2O before combining into four 50 mL tubes and adjusting to 50 mL in each one using ice

cold H_2O. This process was repeated twice more, each time halving the number of tubes in which cells were combined into. Finally, one tube of 50 mL was pelleted at 4,000×g for 20 min at 4 °C before resuspension in 4.5 mL ice cold 10 % (v/v) glycerol. 80 µL aliquots were dispensed into pre-chilled sterile 1.5 mL tubes and flash frozen in LiN_2.

2.5. Plant transformation

2.5.1. Transient transformation of *N. benthamiana*

N. benthamiana plants between 4 and 5 weeks old were transiently transformed by injecting *A. tumefaciens* in agroinfiltration buffer (10 mM MES monohydrate, 10 mM $MgCl_2$, 100 µM acetosyringone, pH 5.7) at OD_{600} of 0.5 or 1.0 into the underside of mature leaves using a 1 mL blunt-ended syringe. To increase expression levels, constructs were co-infiltrated with the RNA silencing suppressor P19-carrying strain (Emma Watts, University of Oxford, UK) (*187*) in a 1:1 ratio.

During co-infiltration experiments, *A. tumefaciens* strains were mixed in equal ratios and at the same OD_{600} in agroinfiltration buffer. In some cases, to ensure adequate expression levels, plants were re-infiltrated 5 d after the first infiltration event with an identical *A. tumefaciens* solution.

2.5.2. Stable transformation of *M. polymorpha*

The spores of *M. polymorpha* were transformed using a method based on published work (*35*). Frozen spores were allowed to thaw at room temperature. For sterilisation, 500 µL 0.1 % (w/v) Sodium dichloroisocyanurate (NaDCC) was added to each tube and

spores resuspended by vortexing or by using a pipette tip to press the sporangia against the wall of the tube. The resuspended spores were left for 5 min before centrifuging for 90 sec at 18,000×g. Most of the sterilisation solution was removed and the spores were resuspended in 500 µL sterile distilled H_2O. The spore suspension was added to 25 mL M51C media (19.8 mM KNO_3, 5 mM NH_4NO_3, 1.5 mM $MgSO_4 \cdot 7H_2O$, 2 mM $CaCl_2 \cdot 2H_2O$, 2 mM KH_2PO_4, 109 µM EDTA-NaFe(III), 45.2 µM KI, 5.55 mM inositol, 81.2 µM nicotinic acid, 48.6 µM pyridoxine-HCl, 296 µM thiamine-HCl, 10 µM $NaMoO_4 \cdot 2H_2O$, 1 µM $CuSO_4 \cdot 5H_2O$, 1.1 µM $CoCl_2 \cdot 6H_2O$, 70 µM $ZnSO_4 \cdot 7H_2O$, 448 µM $MnSO_4 \cdot 4H_2O$, 485 µM H_3BO_3, 2 % [w/v] sucrose, 0.03 % [w/v] L-glutamine, 0.1 % [w/v] casamino acids, pH 5.5) in a 100 mL flask. Spores were grown in growth cabinets with constant light at 23 °C with shaking at 120 RPM for 7 d before transformation.

A single colony of the *Agrobacterium* strain used for transformation was picked into 5 mL LB with appropriate antibiotics and grown overnight at 28 °C with shaking. The bacterial culture was centrifuged at 2,000×g for 15 min, and the pellet resuspended in 10 mL M51C medium supplemented with 100 µM acetosyringone. The *Agrobacterium* was grown for 4-6 h at 28 °C with shaking; if bacteria started to clump the pre-induction period was stopped and 1 mL of the culture added to each flask of 7-day old liquid spore culture. Acetosyringone was added to a final concentration of 100 µM and bacteria and spores co-cultivated for 2 d on a shaker at 120 RPM in the standard growth room already described.

Before plating, spore cultures were poured through a 70 µm cell strainer and immobilized spores washed with at least 150 mL sterile distilled H_2O. A Pasteur pipette was used to move the spores from the strainer to plates containing ½ Gamborg B5

media and appropriate antibiotics. Antibiotic resistant transformants were visible after 2-3 weeks, at which point they were transferred to fresh plates. To generate stable lines gemmae were collected from a single gemmae cup to ensure isogenicity, grown to maturation, and gemmae collected from these plants used in experiments.

For super-transformation of transgenic *M. polymorpha*, a thallus transformation method was used (*34*). Gemmae were plated and grown for 2-3 weeks on ½ MS media in constant light, before cutting of each thallus to remove the apical pieces. The basal fragments were grown under constant light for 3-6 d on ½ MS media to stimulate regeneration, before being added into 50 mL M51C medium containing 100 µM acetosyringone in a 100 mL flask. 1 mL of *Agrobacterium* suspension was added to the flask, which was prepared as follows: a single colony of the strain of interest was picked into 5 mL LB with appropriate antibiotics and grown for 2 d at 28 °C with shaking. The saturated culture was centrifuged for 5 min at 4,000×g and the pellet resuspended in 25 mL M51C with 100 µM acetosyringone. The bacteria and thallus pieces were co-cultured for 3 d under continuous light at 22 °C with shaking at 130 RPM.

The cultures were poured through a 100 µm cell strainer and the thallus pieces washed with at least 500 mL sterile distilled H_2O before being placed in a 50 mL tube containing sterile distilled H_2O and 1 mg mL^{-1} cefotaxime. Thallus pieces were incubated at room temperature for 30 min before three rounds of 20 s vortexing with water changes in between. Thallus fragments were moved onto plates containing ½ MS media and appropriate antibiotics and grown under standard conditions in a growth room as previously described. Transformants were visible after 3-5 weeks. To generate stable lines, gemmae were collected from a single gemmae cup and the plant grown to

maturation, and gemmae collected from these plants were used for experiments.

For selection of transgenic plants, hygromycin (10 μg mL^{-1}) or chlorsulfuron (0.5 μM, from 50 mM stock in DMSO) were used, and cefotaxime (100 μg mL^{-1}) was used to kill any remaining *Agrobacterium* from transformation and to protect against general bacterial contamination.

2.5.3. Extraction and purification of genomic DNA from *M. polymorpha* tissue

Genomic DNA was extracted from *M. polymorpha* using the cetyltrimethylammonium bromide (CTAB) method (*188*). A piece of thallus between 4-8 weeks old and approximately 1 × 1 cm was flash frozen in a 2 mL tube with a metal bead and ground in a TissueLyser Reaction Tube Holder cooled to -80 °C using a TissueLyser II (Qiagen, Hilden, Germany) for 5 min at 25 Hz. Following homogenisation 500 μL of CTAB buffer (2 % CTAB [w/v], 100 mM Tris HCl pH 8, 1.4 M NaCl, 20 mM EDTA pH 8, 1 % [v/v] β-mercaptoethanol) was added, the tube vortexed, and incubated at room temperature for 15 min. 250 μL chloroform: isoamylalcohol (24:1) was added, well mixed, and the mixture centrifuged for 1 min at 18,000×g. The upper phase was taken and 250 μL chloroform: isoamylalcohol (24:1) added again followed by centrifugation in the same way. The upper phase was taken and 500 μL isopropanol added. Centrifugation for 10 min at 18,000×g resulted in pelleting of the DNA; the pellet was then washed with 500 μL 70 % (v/v) ethanol and the tube centrifuged at 20,000×g for 5 min. The ethanol was removed, and the pellet dried at room temperature for approximately 1 h before resuspension in 100 μL distilled water. Extracted genomic DNA was diluted to 20-30 ng

μL^{-1} and used as a template in a standard PCR reaction.

2.6. Plant image analysis

2.6.1. Macroscopic growth analysis of *M. polymorpha*

To quantify plant growth plates containing multiple plants from the same line were scanned at regular intervals using an Epson Perfection V750 PRO scanner (Epson, Hemel Hempstead, UK). Images were analysed using ImageJ (version 1.54, U.S. National Institutes of Health, Bethesda, USA) (*189*). Briefly, colour thresholding was used to detect pixels corresponding to plants in the images and the resulting areas quantified using the Analyze Particles function. A ruler was included in each image and used to set the scale of each image, which allowed for conversion of area from pi^2 to mm^2.

For high resolution images of individual plants images were taken using a Stemi 508 stereomicroscope (Zeiss).

2.6.2. Tissue fixing and staining with Nile Blue A

The staining of PHB granules was based on published work (*51*). Briefly, whole *M. polymorpha* gemmae were fixed in 3 % (v/v) glutaraldehyde in 0.1 M phosphate buffer (93.5 mM K_2HPO_4, 6.52 mM KH_2PO_4, pH 8.0) for 3 h with shaking. The gemmae were then rinsed in water and stained for 5 min in 1 % (w/v) Nile Blue A (Sigma, St Louis, Missouri, USA) before rinsing three times in excess water. Finally, the tissue was soaked for 1 min in 8 % (v/v) acetic acid, rinsed a final time in water, and dry mounted onto a microscope slide.

2.6.3. Confocal laser scanning microscopy

Leaf epidermal samples of *N. benthamiana* and fixed *M. polymorpha* gemmae stained using Nile Blue A were imaged using a Leica SP5 confocal microscope (Leica Microsystems, Wetzlar, Germany). Leaf disks were excised from live leaves and mounted with the abaxial side upwards in perfluorodecalin (Merck, Darmstadt, Germany) whereas fixed and stained gemmae were dry mounted as described. Images were captured with an HC PL APO 20× water immersion objective lens or an HC PL APO CS2 63× water immersion objective lens (both Leica Microsystems, Wetzlar, Germany) at 512×512 px resolution with 4-line averaging and gain adjusted to maximize signal whilst avoiding saturation. Saturation was evaluated using the LAS AF Application Software (Leica Microsystems, Wetzlar, Germany) prior to image acquisition. An excitation wavelength of 488 nm for mClover3 or 594 nm for mCherry was used, with emission collected at 510-540 nm or 600-630 nm, respectively. For Nile Blue A an excitation wavelength of 633 nm was used with emission collected at 650-700 nm. When collecting z-stacks 1-line averaging was used to reduce the delay between images.

For the nuclear staining and fluorescence recovery after photobleaching (FRAP) experiments, samples were imaged using a Zeiss LSM 880 confocal microscope with a LD LCI Plan-Apochromat 25x/0.8 mm water immersion objective lens (Carl Zeiss AG, Oberkochen, Germany). To stain nuclei the underside of agroinfiltrated *N. benthamiana* leaves were injected with 100 ng mL^{-1} 4',6-diamidino-2-phenylindole (DAPI) in phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) using a 1 mL blunt-ended syringe. Leaf samples were excised and

mounted as previously described and images captured after 5 min with excitation at 405 nm (DAPI), 488 nm (mClover3), and 561 nm (mCherry), and emission collected at 430-460 nm (DAPI), 510-540 nm (mClover3), and 600-630 nm (mCherry). For FRAP experiments, circular regions encompassing an entire or part of a compartment were bleached using the 488 nm laser at 100 % intensity for 100 iterations. The minimum size of bleached area was limited by the size of the laser. Images were taken for five pre- and 70 post-bleach scans with an excitation of 488 nm and emission collected at 510-540 nm with the delay between timeseries images minimized. Images were kept for analysis if signal was reduced by at least 50 % in the bleached area, and laser intensity during imaging was decreased to avoid any photobleaching as revealed by monitoring signal in a non-bleached area.

2.6.4. Confocal image analysis

To calculate condensate size and number, a z-stack image spanning a single cell was captured on a Leica SP5 confocal microscope (Leica Microsystems, Wetzlar, Germany) using the settings for mClover3 presented in 2.6.3 and was processed using ImageJ and the 3D Objects Counter plugin package (version 1.5.1;(*190*)). The threshold for object detection and size was manually adjusted for each image based on visual inspection to detect all condensates. The analysis returned a list of detected condensates and their volume in μm^3 in each image.

FRAP data was analysed using ImageJ and the Stowers ImageJ Plugins (version 2: Jay Unruh, Stowers Institute for Medical Research, Kansas City, Missouri, USA). Briefly, a circular bleached area was selected on a time course image and average pixel

intensity quantified over time. For comparison across images pixel intensity was normalised by setting the maximum of each recovery curve to 1 and the minimum to 0. To calculate the average diffusion coefficient from FRAP experiments, normalised recovery curves were fitted to the following function:

$$f(t) = A(1 - e^{\frac{-t}{\tau}})$$

The diffusion coefficient (D) was then estimated as $D \approx \frac{r^2}{\tau}$ where r is the radius of the bleached area and τ the recovery timescale (*191*). τ was converted to units of s using the frame speed of each image.

2.7. Protein purification and analysis

2.7.1. Detection and quantification of luciferase expression

The PULSE system was tested for activity in *Marchantia* by the addition of exogenous D-luciferin solution (1 mM D-luciferin, 0.01 % [v/v] Triton X-100) using a spray bottle to plants grown on plates. Plants were placed in red light (66 µmol photons m^{-2} s^{-1} of 660 nm monochromatic red light [Niello]) for 4 h at 22 °C. The resulting luminescence was visualized using a NightOWL photon counting camera (Berthold Technologies, Bad Wildbad, Germany) with an exposure time of 300 s. This approach was also used to screen for successful PULSE transformants.

Luciferase expression dynamics were determined using a NightSHADE machine (Berthold Technologies, Bad Wildbad, Germany) allowing for automated imaging and light control. One gemma was plated onto 1.5 mL ½ Gamborg B5 media in each well of

a clear 24-well plate, the plate was sealed with parafilm, and the plants grown under normal conditions for 15 d. Prior to placing in the NightSHADE the lid was removed, 150 µL D-luciferin solution was added to each well, and the plate was sealed with optically clear film. Eight small holes were made in the film around the edges of the well to prevent condensation. Plates were placed in the NightSHADE at 22 °C and illuminated with white light (10 µmol photons m^{-2} s^{-1} each of red [660 nm], blue [470 nm], and far-red [730 nm] light) for 12 h followed by 10 µmol photons m^{-2} s^{-1} monochromatic red light for 12 h, and lastly 24 h in white light as described. Image acquisition was automated using the NightSHADE with images taken every 2 h during the entire experimental period, with an exposure time of 300 s. Fresh weight of each plant was measured at the end of the experimental period and used to normalize the photon counts for each plant which were obtained using the WinLight software (Berthold Technologies, Bad Wildbad, Germany).

Luciferase expression was quantified using a DualGlo Luciferase Assay (Promega, Madison, Wisconsin, USA). Tissue was ground in a TissueLyser Reaction Tube Holder cooled to -80 °C using a TissueLyser II (Qiagen, Hilden, Germany) for 5 min at 25 Hz, spun down briefly, and 200 µL Passive Lysis Buffer (PLB; Promega, Madison, Wisconsin, USA) was added. Samples were vortexed to resuspend the pelleted tissue and centrifuged for 2 min at 18,000×g. 75 µL of each sample was added to each well followed by 75 µL DualGlo Luciferase reagent (Promega, Madison, Wisconsin, USA). After waiting 15 min, firefly luminescence was measured using a GloMax Plate Reader (Promega, Madison, Wisconsin, USA). To measure *Renilla* luciferase activity, 75 µL of DualGlo Stop&Glo reagent (Promega, Madison, Wisconsin, USA) was added to each

well and luminescence measured in the same way after 15 min. To allow comparisons across experiments a standard curve using pure firefly luciferase standard (Promega, Madison, Wisconsin, USA) was used.

2.7.2. Purification of overexpressed recombinant proteins from plant material

Mature leaf samples from *N. benthamiana* and thallus fragments from 4–8-week-old *M. polymorpha* were collected and weighed before flash freezing with a metal bead in LiN_2. Samples were homogenised in a TissueLyser Reaction Tube Holder cooled to -80 °C using a TissueLyser II (Qiagen, Hilden, Germany) for 5 min at 25 Hz and resuspended in a 10-fold excess (volume to weight) of cold grinding buffer (50 mM NaH_2PO_4, 150 mM NaCl, 50 mM ascorbic acid, 0.6 % [w/v] PVPP-40, 0.4 % [w/v] bovine serum albumin, 5 % [v/v] glycerol, 1 % [v/v] TWEEN-20, one mini tablet of Pierce Protease Inhibitor [Thermo Fisher Scientific, Waltham, USA], pH 8.0). The lysate was sonicated for 10 min using a Grant Ultrasonic Bath XB3 (Keison Products, Chelmsford, UK) and centrifuged for 15 min at 3,000×g at 4 °C. The supernatant was collected and kept on ice, and the pellet resuspended in 2 mL cold grinding buffer and centrifuged as above; the supernatant was then added to the already collected sample.

For Ni-affinity purification of His-tagged proteins, the collected supernatant was added to a 20 mL Econo-Pac gravity-flow chromatography column (Bio-Rad, Hercules, USA) containing 4 mL of Ni-NTA Agarose resin (Thermo Fisher Scientific, Waltham, USA) and allowed to bind for 18 h on a rotating shaker at 4 °C. The column was then washed three times with 10 mL native wash buffer (50 mM NaH_2PO_4, 150 mM NaCl, 20

mM imidazole, 5 % [v/v] glycerol, 1 % [v/v] TWEEN-20, pH 8.0) and eluted using three washes of 3 mL native elution buffer (50 mM NaH_2PO_4, 500 mM NaCl, 250 mM imidazole, 5 % [v/v] glycerol, 0.1 % [v/v] TWEEN-20, one mini tablet of Pierce Protease Inhibitor [Thermo Fisher Scientific, Waltham, USA], pH 8.0).

2.7.3. Analysis of proteins by sodium-dodecyl-sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were mixed with SDS gel-loading buffer (250 mM Tris-HCl [pH 6.8], 8 % [w/v] SDS, 0.1 % [w/v] bromophenol blue, 40 % [v/v] glycerol, 100 mM dithiothreitol) and either boiled at 95 °C for 5 min or placed at 37 °C for 30 min. 5-12 % polyacrylamide gels were prepared by layering a resolving gel (0.375 M Tris HCl pH 8.8, 0.1 % SDS [w/v], 12 % [v/v] acrylamide, 0.05 % w/v] ammonium persulphate [APS], 0.05 % [v/v] N,N,N',N'-tetramethylethylene-diamine [TEMED]) and a stacking gel (0.375 M Tris-HCl pH 8.8, 0.1 % [w/v] SDS, 5 % [v/v] acrylamide, 0.125 % [w/v] APS, 0.125 % [v/v] TEMED). 10 – 15 µL sample was loaded into each well. Electrophoresis was done in Tris-Glycine-SDS buffer (25 mM Tris base, 200 mM glycine, 3.5 mM SDS) at a constant voltage of 120 V until the dye front reached the bottom of the gel. To visualize protein bands the gel was stained in ReadyBlue Protein Gel Stain (Merck, Darmstadt, Germany) for 10 min, rinsed in distilled H_2O, and imaged using an ImageQuant LAS4000 camera system (GE Healthcare, Illinois, USA).

2.7.4. Western blotting

After SDS-PAGE protein bands were transferred to a nitrocellulose membrane using electroblotting (100 V, 1 h) in cold transfer buffer (20 % [v/v] methanol, 50 mM Tris base,

250 mM glycine, 3.5 mM SDS). The membrane was blocked using bovine milk albumin (10 mM Tris-HCl [pH 8], 150 mM NaCl, 0.5 % [v/v] TWEEN-20, 5 % [w/v] skimmed milk powder) with shaking at room temperature for 1 h and was then incubated with the primary antibody in skimmed milk solution at 4 °C overnight with shaking. To detect mClover3 and mCherry, α-GFP and α-mCherry antibodies (both Proteintech Europe, Manchester, UK) were used, respectively. To detect β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase, α-Myc (Abcam, Cambridge, United Kingdom), α-PhbB (kindly donated by Yves Poirier, Université de Lausanne, Switzerland) (*56*), and α-HA (Merck, Darmstadt, Germany) antibodies were used, respectively. Commercial antibodies were used at 1 µg mL^{-1} and the α-PhbB antibody was used at a dilution of 1:500. Membranes were washed three times (10 mM Tris HCl pH 8, 150 mM NaCl, 0.5 % [v/v] TWEEN-20) for 5 min each and were then incubated in all cases with a goat α-rabbit IgG antibody linked to horseradish peroxidase (HRP) (Merck, Darmstadt, Germany) at 0.1 µg mL^{-1} for 1 h at room temperature with shaking. Membranes were washed three times as described before treatment for 2 min with 0.1 mL cm^{-2} EZ-ECL Chemiluminescence Detection Kit for HRP (Biological Industries, Kibbutz Beit-Haemek, Israel), composed of a 1:1 ratio of peroxide solution and luminol. HRP signal was visualized using an ImageQuant LAS4000 camera system (GE Healthcare, Illinois, USA).

2.8. Polymer and metabolite analysis

2.8.1. PHB extraction

The extraction of PHB was based on a method outlined in (*52*). 20-250 mg of plant

tissue was collected and flash frozen with a metal bead in LiN_2. The tissue was homogenised in a TissueLyser Reaction Tube Holder cooled to -80 °C using a TissueLyser II (Qiagen, Hilden, Germany) for 5 min at 25 Hz, transferred to a glass tube and extracted in 4 mL methanol for 1 h at 55 °C and subsequently extracted twice in 4 mL ethanol for 1 h at 55 °C. This was followed by the addition of 500 µL chloroform, 1.7 mL ethanol, 0.2 mL HCl, and 2 µL internal standard 3-hydroxy-valerate-methyl-ester to each sample, which was incubated for 4 h at 100 °C, resulting in the formation of 3-hydroxybutyrate ethyl ester. Lastly 4 mL of 0.9 M NaCl was added, the tubes vortexed, and the phases separated by centrifugation at 4,000×g for 10 min at 4 °C. The upper aqueous phase was removed and 25 µL of the lower phase was derivatised with 75 µL of N-methyl N-trimethylsilyltrifluoroacetamide (MSTFA) for 30 min at room temperature with gentle shaking.

2.8.2. Citramalate and general polar metabolite extraction

For citramalate extraction and general polar metabolite analysis, a method developed by Lisec et al was used (*192*). 20 – 200 mg of plant tissue was flash frozen with a metal bead in LiN_2 and homogenised in a TissueLyser Reaction Tube Holder cooled to -80 °C using a TissueLyser II (Qiagen, Hilden, Germany) for 2 min at 30 Hz. After homogenisation 400 µL pre-cooled methanol was added, the tube was vortexed, followed by addition of 60 µL 0.1 mg mL^{-1} norvaline as the internal standard. Samples were incubated at 70 °C for 15 min with shaking, before being centrifuged for 10 min at 20,000×g. The supernatant was transferred to another tube and 250 µL chloroform and 500 µL water added. Each tube was vortexed for 15 sec and then centrifuged for 15 min at 1700×g. Two samples of 100 µL were taken from each tube to generate duplicate

samples to be dried down under vacuum. To derivatise the dried samples, 25 µL pyridine was added and the samples incubated at 37 °C with shaking at 950 RPM for 90 min followed by the addition of 35 µL *tert*-butyldimethylsilane (tBDMS) and incubation at 65 °C with shaking at 950 RPM for 1 h.

2.8.3. Gas chromatography-mass spectrometry (GC-MS)

0.25 µL (PHB quantification), 0.5 µL (citramalate quantification), or 1 µL (polar metabolite analysis) was injected in splitless mode at 230 °C into an Intuvo 900 gas chromatography mass spectrometer (Agilent Technologies, Santa Clara, USA) fitted with a 15 m × 250 µm × 0.25 µm DB-5ms Ultra Inert Intuvo GC capillary column module (Agilent Technologies, Santa Clara, USA). For PHB quantification, chromatography was achieved by holding the temperature at 60 °C for 5 min followed by heating at a rate of 7.5 °C min^{-1} to 200 °C resulting in a run time of 24 min. For the quantification of citramalate the temperature was held at 90 °C for 3 min followed by heating at a rate of 7.5 °C min^{-1} to 250 °C resulting in a run time of 24 min. Finally, for the general polar metabolite analysis, temperature was held at 100 °C for 5 min followed by heating at a rate of 5 °C min^{-1} to 300 °C resulting in a run time of 45 min. For the general polar metabolite analysis, an Agilent 5977B mass spectrometer (Agilent Technologies, Santa Clara, USA) was used in scanning mode with a solvent delay of 7.5 min and a scanning window of 50 – 600 *m/z*. For both PHB and citramalate quantification, the same spectrometer was used in single ion monitoring (SIM) mode to detect the 189 and 203 *m/z* fragment ions for PHB and 3-hydroxy-valerate-methyl-ester respectively, and the 433 and 288 *m/z* fragment ions for citramalate and norvaline quantification respectively. Samples were injected in triplicate and average values taken forward for

analysis. In all cases the MassHunter MS Quantitative Analysis software (Agilent Technologies, Santa Clara, USA) was used for peak identification and integration with the help of the National Institute of Standards and Technology (NIST, Gaithersburg, USA) mass spectrum library (version 2.0). PHB:3-hydroxy-valerate-methyl-ester and citramalate:norvaline standard curves were generated using pure PHB and citramalate standards (Merck, Darmstadt, Germany) and used for absolute quantification. Standard curves for the general polar metabolite analysis were prepared and run by Pedro Bota (University of Oxford, UK). A mixture of 21 organic acids and amino acids (all Merck, Darmstadt, Germany) at 0.1 mg mL^{-1} was prepared and used to generate standard curves with norvaline as the internal standard.

3. Construction and analysis of a computational model to identify strategies for optimising the biosynthesis of PHB in *M. polymorpha*

3.1. Introduction

The biosynthesis of PHB has been engineered into a range of organisms, including bacteria, yeast, and plants. In most cases this has involved only the introduction of the three enzymes directly responsible for PHB biosynthesis from *R. eutropha*, namely β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase (Figure 1.1A). Very few examples exist where the engineering of PHB has been accompanied by a consideration of the impact on the wider metabolic network, or whether specific changes to the capacity of the existing metabolic network might be needed to support maximal PHB production. As will be discussed in section 3.7.2, the only such 'ancillary target' that has been previously considered is ACL activity (*61, 193, 194*). In addition, an example exists of the manipulation of photosynthesis and plant growth alongside PHB biosynthesis in switchgrass (*47, 58*). However, these are trial-and-error-based approaches and do not include a systematic analysis and understanding of impacts across the entire metabolic network.

A whole-network analysis is necessary given the complexity and interconnectedness of metabolism; the behaviour of the metabolic network might be non-intuitive and many examples exist of metabolic interventions resulting in different or opposite effects to those expected (*9*). For this reason, computational models of the whole metabolic network have become important tools for the design of metabolic engineering strategies. Metabolic modelling usually takes one of two approaches; the first, kinetic modelling, relies on knowledge regarding reaction stoichiometry, rate equation, and kinetic parameters of each enzyme in the system. Kinetic modelling usually involves the integration of a large system of non-linear ordinary differential equations; this poses limits in terms of computation time and ultimately limits the size of the network which can be modelled. As such only subsets of plant metabolism have currently been modelled using this approach, such as photosynthesis (*195*).

Stoichiometric models can be built from genome sequence data by extracting knowledge about the number and stoichiometry of metabolic reactions encoded in the genome. By applying a pseudo-steady state assumption and constraining some reactions based on experimental data, the possible flux solution space can be reduced, and linear programming used to find a solution which satisfies an objective using FBA. This approach can model large networks and has been shown capable of making accurate quantitative predictions of plant metabolic behaviour (*196*). In addition, stoichiometric models can be used for *in silico* testing of metabolic engineering such as introduction of heterologous enzymes and knockout or upregulation of target enzymes. Stoichiometric models have been used extensively to inform the engineering of microorganisms (*197*); to my knowledge no example of using

FBA to inform engineering of plants exist.

Given the lack of kinetic knowledge about *M. polymorpha* enzymes, and my desire to model central metabolism in its entirety, a stoichiometric model of the central metabolism of *M. polymorpha* was built and interrogated for its ability to support PHB production. The model was used to establish the optimal subcellular location for PHB production, and to identify changes to the general metabolic network which might be conducive to maximal PHB accumulation. These results were used to inform the design of an engineering strategy implemented experimentally, the results of which are presented in Chapter 5.

3.2. Model construction and testing

A stoichiometric model of the core metabolism of *M. polymorpha* was built as explained in Materials and Methods section 2.1, hereafter referred to as the baseline model. A visual representation of the baseline model is shown in Figure 3.1.

To ensure that the model reflected known plant metabolic behaviour, the model was set up with biomass accumulation as the objective function (i.e. the reaction whose flux should be maximized) and fluxes through central metabolism investigated following optimisation using pFBA. pFBA uses minimization of total flux as a secondary objective; this is based on the principle that evolutionary selective pressures will have reduced the total necessary amount of enzymes to a minimum for efficient use of nitrogen resources (*198*). As such a solution with a lower sum of fluxes represents a flux distribution with higher nitrogen use efficiency.

The model showed a number of flux patterns that are typical of photosynthetic plant tissues over a day-night cycle, indicating that the model behaviour was a reasonable prediction of expected metabolism (Figure 3.2A). For example, the TCA cycle operated differently at day and at night; nocturnal flux was fully cyclical but an incomplete TCA cycle ran during the day (*150*). Other recognised metabolic behaviours included the accumulation of starch and sucrose during the day for use at night, and the accumulation of citrate in the vacuole at night for use during the day to provide carbon skeletons for nitrogen assimilation (*199*). The latter was predicted by the model even without an experimentally derived constraint on nocturnal citrate accumulation rate.

The effects of some of the constraints imposed on the model were investigated in more detail. The constraint on daytime starch accumulation rate, for instance, had a large effect on the provision of nocturnal carbon (Figure 3.2B). The daytime starch accumulation rate was constrained based on starch content of *Marchantia* biomass (*147, 148*) (Table S2.1). When unconstrained the flux through the starch accumulation reaction was ~100-fold higher than the constrained value, and the daytime sucrose accumulation rate decreased by 98 %. As such the model favours starch over sucrose as a nocturnal carbon source. This can be explained given that the export of sucrose from the vacuole at night requires proton symport. Protons have to be pumped into the vacuole using a V-type ATPase which is energetically costly. Perhaps because of this, unconstraining daytime starch accumulation and thereby allowing the model to store more starch resulted in a slightly lower sum of fluxes (i.e. it is slightly more efficient) (Table 3.1).

Constraints were imposed on vacuolar organic acid storage accumulation reactions

based on experimental data showing that *M. polymorpha* accumulates citrate but not fumarate and malate during the night (Corinna Hartinger, University of Oxford, unpublished). Given that the lack of fumarate or malate accumulation in the vacuole is unusual for the photosynthetic tissues of plants (*200*), the effect of these constraints on the fluxes through central metabolism were investigated (Figure 3.2C).

Removing the constraints on the vacuolar accumulation of fumarate and malate led to relatively high daytime fumarate storage fluxes and nocturnal malate storage fluxes, with a small effect on biomass accumulation. It also led to an overall reduction in the sum of fluxes of the model (Table 3.1). One important change in flux concerned the export of citrate from the mitochondrion during the night. In contrast to the baseline model, where most exported citrate was converted to isocitrate in the cytosol and then re-imported into the mitochondrion, all the citrate exported (bar the contribution of citrate to biomass) was translocated to the vacuole. In addition, although fumarate was stored during the day for use at night, all this stored fumarate was converted to malate in the cytosol at night. Similarly, the fumarate that was transported to the vacuole during the day was derived primarily from malate. It is unlikely that this cyclic flux mode reflects *in vivo* metabolism and is instead likely a result of the specific constraints or energetic requirements imposed on the model. Finally, the mechanism by which vacuolar nocturnal stored citrate was transported from the vacuole to the cytosol was affected by allowing the vacuolar storage of malate and fumarate. When only citrate was allowed to accumulate, the export of citrate was achieved with the symport of protons, such that three protons were transported from the vacuole to the cytosol for every citrate molecule. In the case of vacuolar malate and fumarate

storage, however, citrate was exported from the vacuole by way of citrate-malate exchange.

Hence the model of *M. polymorpha* as constructed presents realistic behaviour for a day-night photosynthetic system with many recognised and expected metabolic flux patterns. In addition, some of the experimental constraints (especially those concerning diurnal patterns of carbohydrate and carboxylic acid accumulation) have a significant effect on metabolic fluxes. This suggests that there are specific and unique features of *Marchantia* metabolism that are well captured by the model. As such this model was determined to be suitable for use as a predictive tool for the engineering of PHB in *M. polymorpha*.

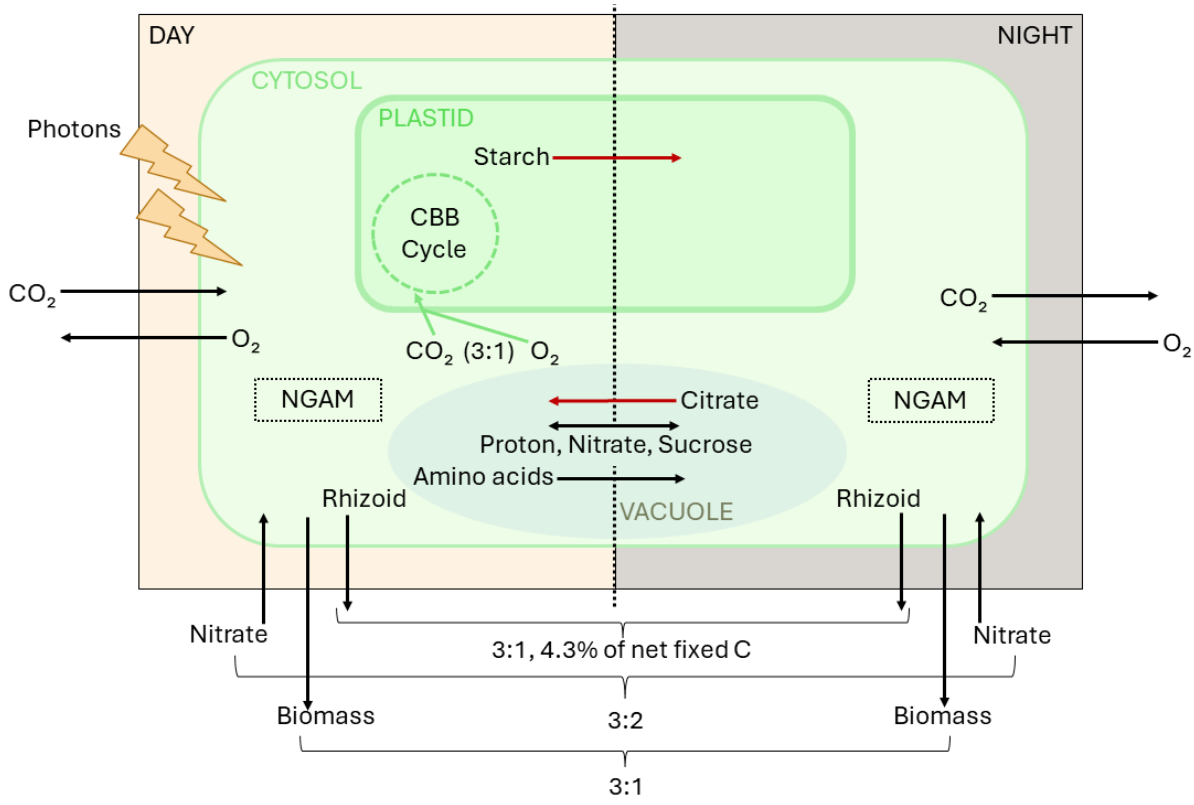
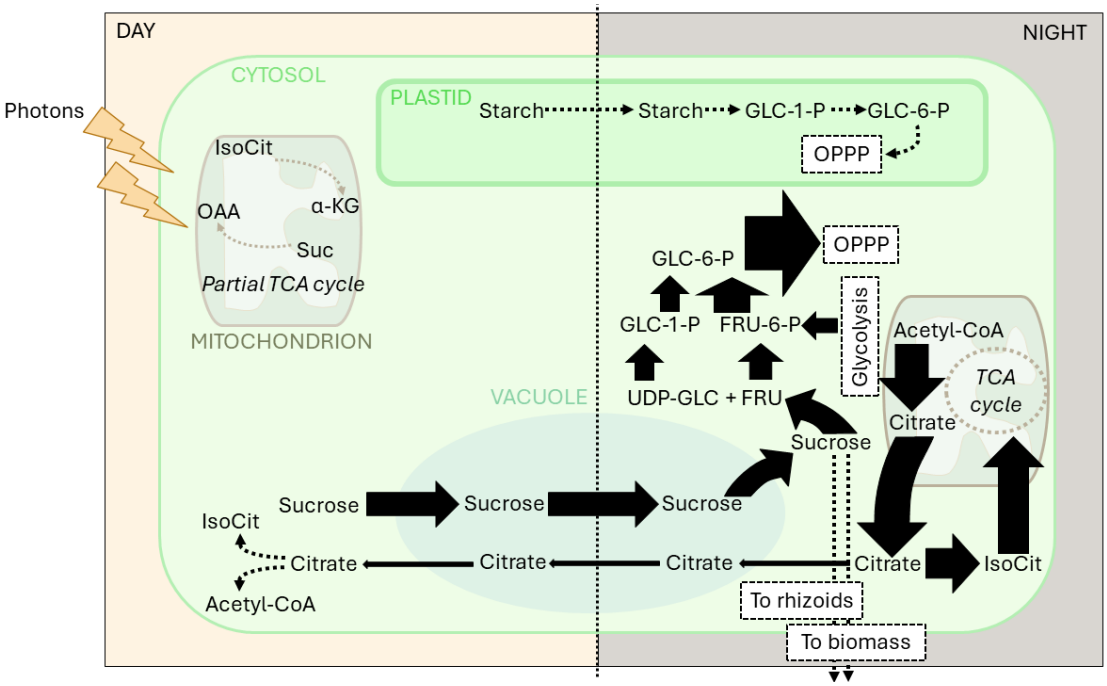


Figure 3.1: Basic schematic of *M. polymorpha* baseline model showing selected inputs, outputs, and constraints. Fluxes constrained to a specific value or to a range of values are represented as red arrows. NGAM, non-growth associated maintenance; C, carbon; CBB, Calvin-Benson-Bassham. For an explanation of all constraints please refer to Table S2.1.

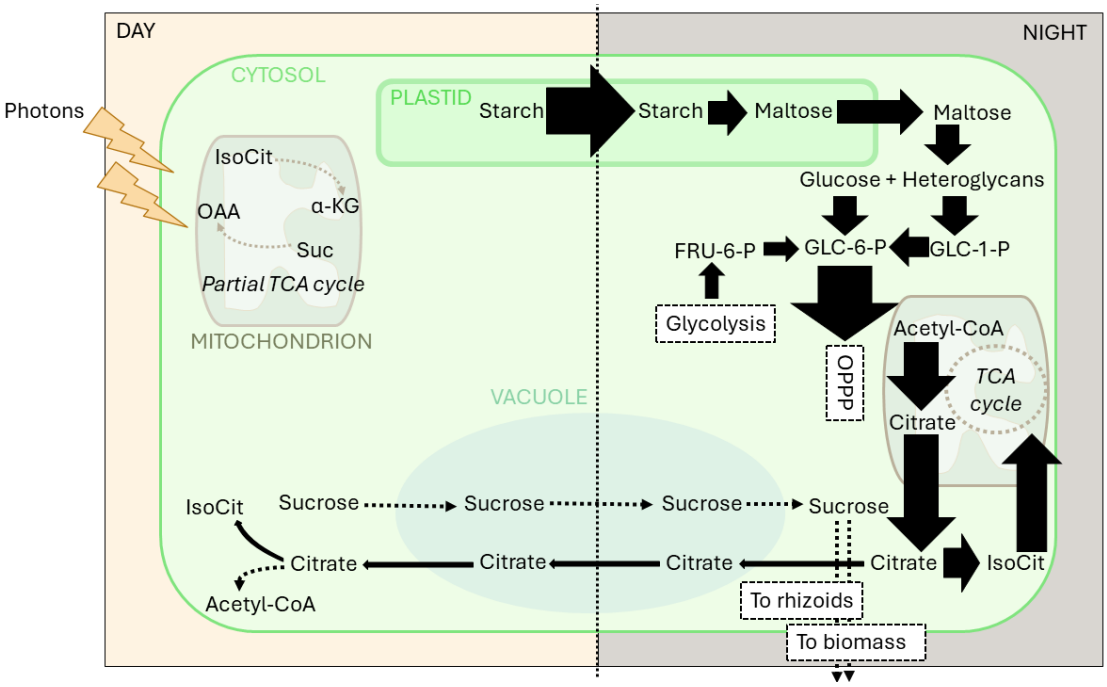
Table 3.1: The flux to biomass accumulation in the baseline model, compared to the baseline model with starch storage or vacuolar malate/fumarate storage unconstrained. All fluxes are in units μmol m^{-2} s^{-1}.

Model	Flux to biomass	Sum of fluxes
Baseline model ...	0.0264	175.2
... with unconstrained starch storage	0.0265 (+ 0.4 %)	175.1
... with unconstrained fumarate and malate storage	0.0259 (- 1.9 %)	173.0

A.


DAY
NIGHT
Photons
CYTOSOL
PLASTID
Starch
Starch
GLC-1-P
GLC-6-P
OPPP
IsoCit
OAA
α-KG
Suc
Partial TCA cycle
MITOCHONDRION
GLC-6-P
OPPP
GLC-1-P
FRU-6-P
Glycolysis
Acetyl-CoA
TCA cycle
Citrate
VACUOLE
UDP-GLC + FRU
Sucrose
Sucrose
Sucrose
Sucrose
IsoCit
Citrate
Citrate
Citrate
Citrate
IsoCit
Acetyl-CoA
To rhizoids
To biomass

B.


DAY
NIGHT
Photons
CYTOSOL
PLASTID
Starch
Starch
Maltose
Maltose
Glucose + Heteroglycans
IsoCit
OAA
α-KG
Suc
Partial TCA cycle
MITOCHONDRION
FRU-6-P
GLC-6-P
GLC-1-P
Glycolysis
Acetyl-CoA
TCA cycle
OPPP
Citrate
VACUOLE
IsoCit
Sucrose
Sucrose
Sucrose
Sucrose
Citrate
Citrate
Citrate
Citrate
IsoCit
Acetyl-CoA
To rhizoids
To biomass

C.

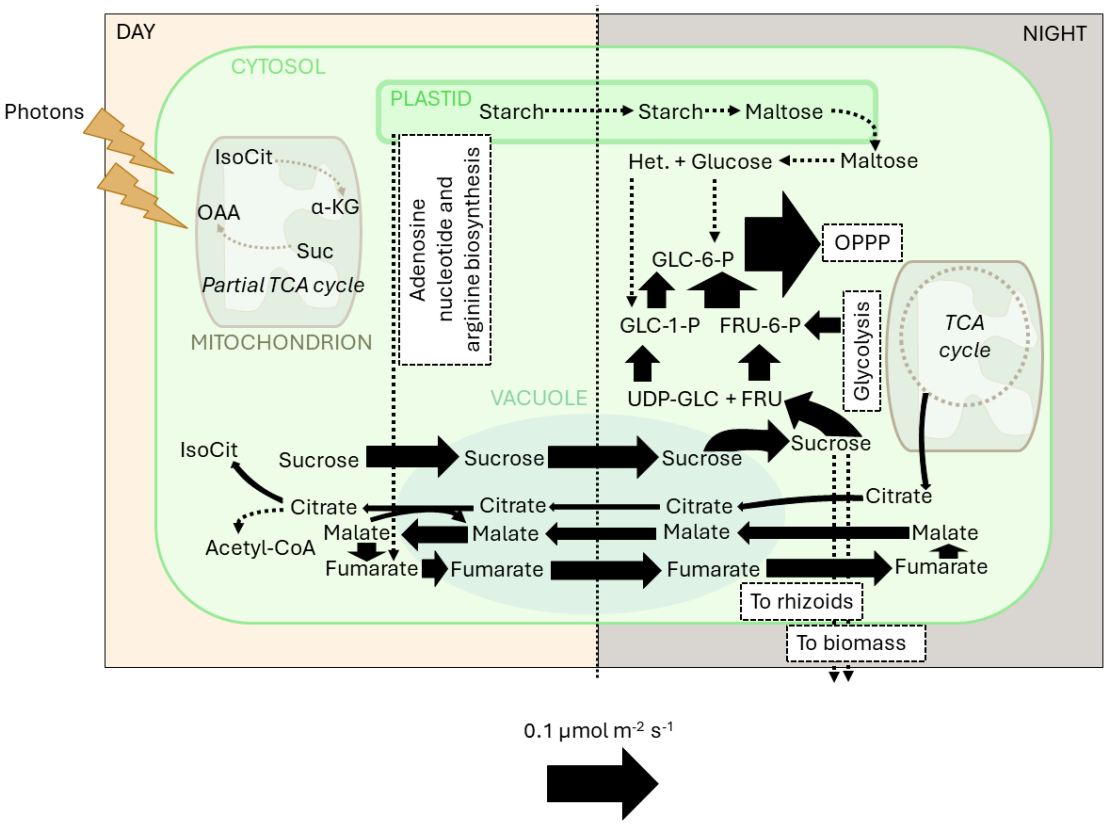


Figure 3.2: Major fluxes through the *M. polymorpha* baseline model with biomass accumulation as the objective. Relevant flux paths are shown for the baseline model (A), and the model with unconstrained starch storage (B) or unconstrained malate and fumarate storage (C). Width of black arrows are scaled according to the flux they carry in units μmol m^{-2} s^{-1}. Fluxes below 0.01 are represented using a dotted line. α-KG, α-ketoglutarate; OAA, oxaloacetate; Suc, succinate; IsoCit, isocitrate; FRU, fructose; UDP-GLC, UDP-glucose; FRU-6-P, fructose-6-phosphate; GLC-1-P, glucose-1-phosphate; GLC-6-P, glucose-6-phosphate; Het., heteroglycans; OPPP, oxidative pentose phosphate pathway.

3.3. Modification of the baseline model for PHB synthesis

PHB biosynthesis has been engineered in the cytosol, plastid, mitochondria, and peroxisomes of plants (*54*). It has become widely accepted that mitochondria are a sub-optimal location for PHB production because of the rapid and efficient use of acetyl-CoA in cellular respiration (*54*). The production of PHB in the peroxisomes,

however, could take advantage of higher levels of acetyl-CoA and reducing power present in this organelle due to its role in fatty acid β-oxidation, although that would require substantial rates of lipid mobilisation to be taking place, such as during oilseed germination. In sugarcane, PHB has been produced in peroxisomes to a relatively modest 1.6 % per unit leaf dry weight. Although a limited study on the effect of PHB production in sugarcane peroxisomes did not reveal any deleterious effects on overall plant health, when produced in the peroxisomes of transgenic *Arabidopsis* a significant reduction in hypocotyl elongation was seen despite accumulation to only 0.1 % per unit total plant dry weight (*201*).

The two best studied subcellular locations for PHB biosynthesis are the cytosol and the plastid; the plastid has resulted in the highest yield of PHB in a plant to date (40 % per unit leaf dry weight), and production in the cytosol results in low to moderate yields (to a maximum of 0.8 % per unit leaf dry weight [converted from value reported as % per unit leaf fresh weight using conversion factor from (*202*)]; Table 1.1). In both cases the production of PHB was accompanied by deleterious effects on plant health; in cytosolic production this is primarily attributed to competition between PHB and biomass accumulation for acetyl-CoA, whereas in the plastid competition effects are thought to be accompanied by issues caused by the accumulation of insoluble PHB granules. In exploring PHB as a case study for the metabolic modelling of *M. polymorpha* I decided to focus on cytosolic and plastidial production; in the case of cytosolic accumulation the availability of acetyl-CoA provided a potentially straightforward engineering target whilst enabling simpler engineering, whereas engineering plastidial PHB production might allow for a record PHB yield if a suitable

engineering strategy presented itself as a result of the modelling.

As such the baseline model was modified to allow for the accumulation of PHB in both the cytosol and the plastid (Table 2.1). Briefly, reactions were added for the conversion of acetyl-CoA to PHB according to the 3-enzyme pathway from *R. eutropha* (Figure 1.1A). In addition, efforts to produce PHB in plants have relied either on the β-ketothiolase encoded by *phbA* or the acetoacetyl-CoA synthase encoded by *NphT7* to catalyse the first step of the process. Whereas β-ketothiolase condenses two molecules of acetyl-CoA to form acetoacetyl-CoA, acetoacetyl-CoA synthase catalyses the condensation of acetyl-CoA and malonyl-CoA with concomitant formation of CoA. Studies have shown that β-ketothiolase is the main contributor to low PHB yield in sugarcane mesophyll cells (*59*), and that the constitutive expression of the *phbA* gene product is deleterious to transformation efficiency in a range of plants (*71*). The low yield in sugarcane cells could be overcome by using acetoacetyl-CoA synthase; for this reason, I wanted to explore both pathways to PHB production and both versions of the first step towards PHB production were added to the model.

As explained in section 2.1.3, I modelled both the day and the night in a single optimization step by the duplication of most reactions and the introduction of linker reactions. As such PHB was allowed to be produced both in the day and the night in the plastid or the cytosol, subsequently referred to as total plastidial or cytosolic PHB, respectively.

PHB production is hypothesized to be detrimental to plant health by directly competing with biomass-accumulation pathways such as fatty acid synthesis for acetyl-CoA. For this reason, I explored the trade-off between PHB production and

biomass accumulation using Pareto analysis (Figure S3.1). The perfect linear relationship between both cytosolic and plastidial PHB and biomass accumulation indicated that the two objectives were in direct competition with one another. For this reason, to force biomass accumulation and PHB production to happen simultaneously, I had to introduce a constraint. It has been reported that PHB production to 10-12.5 % per unit whole plant dry weight is needed for plant-based production to be cost-competitive with petroleum-derived plastics (*46, 47*). To model this situation, a constraint was added to the baseline model such that the flux to either total cytosolic or plastidial PHB accumulation was 10% of the flux to biomass thereby allowing for simultaneous biomass and PHB production. These became the cytosolic PHB and plastidial PHB models, respectively.

pFBA was used to solve for the fluxes through the network given the above constraints. The resulting fluxes to biomass accumulation and PHB production using pFBA with biomass as the objective, as well as the sum of fluxes of the different models, are illustrated in Table 3.2. As expected given the trade-off between biomass and PHB production, both cytosolic and plastidial PHB production resulted in decreased flux to biomass. There was very little difference between the two subcellular locations in terms of flux to PHB and sum of fluxes.

Table 3.2: Flux to biomass, flux to PHB, and sum of fluxes of the baseline model with and without cytosolic or plastidial PHB production. Percentage change compared to the baseline model is indicated. All fluxes are in units μmol m^{-2} s^{-1}.

Model	Flux to biomass	Flux to PHB	Sum of fluxes
Baseline	0.0264	0	175.2
Baseline + 10 % cytosolic PHB	0.0259 (- 1.9 %)	0.00259	175.4 (+ 0.1 %)
Baseline + 10 % plastidial PHB	0.0259 (- 1.9 %)	0.00259	175.2 (± 0 %)

3.4. Optimal predicted paths to plastidial and cytosolic PHB production

The main flux paths to cytosolic and plastidial PHB accumulation as predicted by the model were investigated and are illustrated in Figure 3.3. In both cases PHB was exclusively produced during the day (Figure S3.1). Daytime production schemes are favoured as they result in a higher yield and lower sum of fluxes. This reflects access to higher levels of ATP, reducing-power and carbon-containing precursors during the day as a result of photosynthesis.

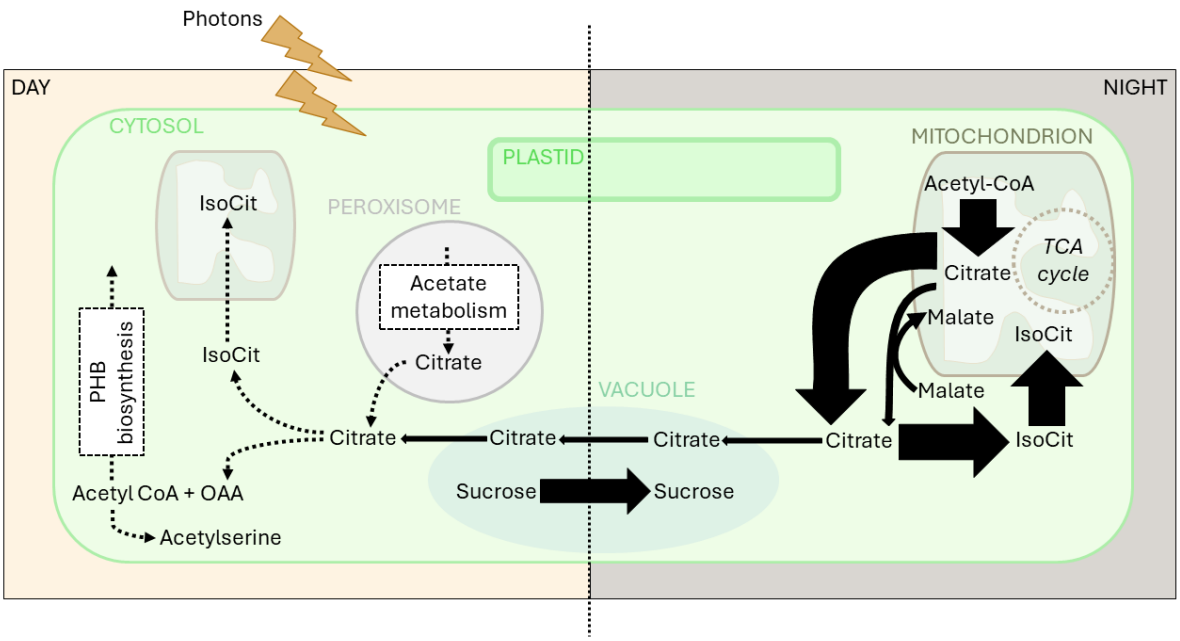
The optimal predicted path to plastidial PHB production involved the export of threonine from the plastid, which was converted to acetaldehyde in the cytosol, and imported into the mitochondrion to be converted to acetate by aldehyde dehydrogenase. The acetate was re-imported into the plastid where it was converted to acetyl-CoA by acetyl-CoA synthetase and used for PHB production. Only 15 % of the plastidial acetyl-CoA pool was converted to PHB. This reflects the wider use of acetyl-CoA in plastidial metabolism; the biggest flux of plastidial acetyl-CoA was towards fatty acid biosynthesis.

An alternative route for the production of acetyl-CoA in the plastid is via the pyruvate dehydrogenase (PDH) complex, which catalyses the oxidation of pyruvate to acetyl-CoA, NADH, and CO_2. *In vivo*, this is thought to be the major source of acetyl-CoA in the plastid (*203*). However, forcing flux through daytime PDH led to an increase in the sum of fluxes by 0.7 %. According to the model, using PDH rather than acetyl-CoA synthetase to produce plastidial PHB is less efficient.

The production of PHB in the cytosol relied on citrate provision; 11.7 % of cytosolic citrate was obtained following export from the peroxisome and the remaining 88.3 % was exported from the vacuole. The vacuolar citrate was the product of the nocturnal mitochondrial TCA cycle and was exported from the mitochondrion by a malate-citrate exchanger and an isocitrate-citrate exchanger. The peroxisomal citrate was produced from acetyl-CoA in the peroxisome, which was itself the product of acetate imported from the cytosol. The cytosolic acetate was produced in the same way as in the plastidial PHB production case explained above, i.e. from mitochondrial acetaldehyde itself derived from threonine biosynthesis in the plastid. The model used cytosolic ACL to produce acetyl-CoA from the cytosolic citrate pool, most of which (92.3 %) went on to get converted to PHB.

Neither cytosolic nor plastidial PHB production made use of the acetoacetyl-CoA synthase encoded by *NphT7*. By setting the flux through β-ketothiolase to 0 I was able to force the model to use acetoacetyl-CoA synthase instead to produce acetoacetyl-CoA. The only change in the resulting flux distribution was the conversion of acetyl-CoA to malonyl-CoA via either the plastidial or cytosolic ACC, which was then used by acetoacetyl-CoA synthase. As such, the versions of the models which used acetoacetyl-CoA synthase carried flux through an extra reaction and thereby had a higher sum of fluxes; using pFBA means this pathway was not chosen.

A.



B.

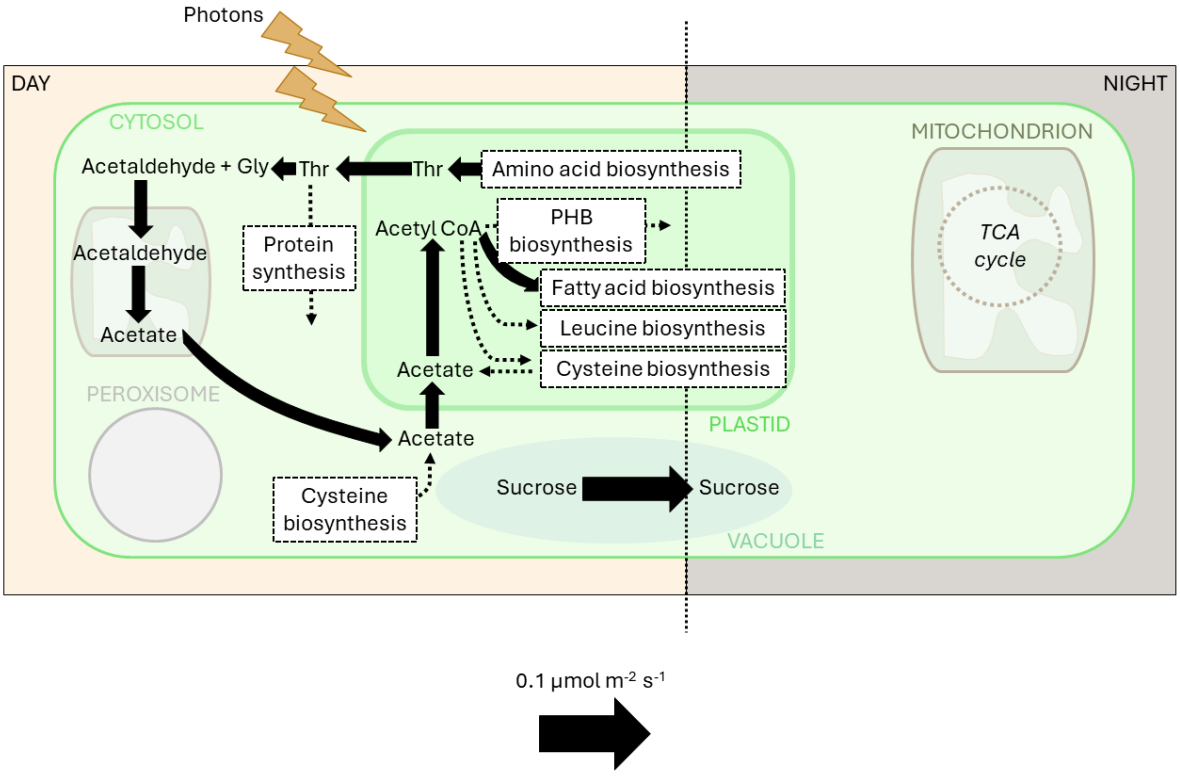


Figure 3.3: An illustration of the predicted optimal flux paths towards cytosolic (A) and plastidial (B) PHB production in *M. polymorpha*. Width of black arrows are scaled according to the flux they carry in units µmol m^{-2} s^{-1}. Fluxes below 0.01 are represented using a dotted line. α-KG, α-ketoglutarate; OAA, oxaloacetate; IsoCit, isocitrate.

One of the limitations of using FBA is that fluxes in an optimal FBA solution are not necessarily uniquely defined. This limitation was addressed by using flux variability analysis (FVA) (*204*); flux distributions were analysed to identify all fluxes in the PHB-producing models that differed from the baseline model. Fluxes were considered different to baseline if the range of possible values for the respective reaction flux in the PHB and baseline models did not overlap. The analysis was extended to reveal the reactions in the PHB-producing models likely to have a significant effect on the flux distribution because (a) the flux ranges of the same reaction differed by a factor of 1.1 or more between PHB-producing and baseline scenarios, (b) the reaction only carried flux in one model, or (c) the flux carried by a reaction entailed a different direction in each model (e.g. import vs export from a compartment) (Table 3.3). A factor of 1.1 (equivalent to a 10 % change) was chosen so that the analysis would return the reactions directly responsible for PHB production. A full list of reactions which satisfied these requirements can be found at https://github.com/AnyaLindstromBattle/Thesis.

Although the cytosolic PHB-producing model had fewer reactions that were different to the baseline model, there were twice as many reactions where this difference was likely to have a significant impact on metabolism according to the definition used in this analysis. Neither model required any reactions to run in a different direction to the baseline model, but both cytosolic and plastidial PHB production relied on the increase or decrease in flux of a significant number of reactions. In the cytosolic production model, a large proportion of these reactions were related to nocturnal TCA cycle activity, responsible for most of the citrate used in PHB biosynthesis (Figure 3.3).

Interestingly, no reaction in the plastidial version had a fold change compared to baseline higher than 1.55. In the cytosolic model, however, 49 % of the reactions 1.1-fold or more different between the two models differed by more than 1.55-fold. The most striking of these was the 12.7-fold increase in the flux through ACL, which underlines the importance of the activity of this enzyme to cytosolic PHB production.

Table 3.3: Result of FVA between the baseline model and the plastidial or cytosolic PHB production models. The number of reactions satisfying the stated conditions are shown for each model. Reactions which always carried a flux < 1^{-7} µmol m^{-2} s^{-1} were ignored.

Condition	Cytosol	Plastid
Reactions with non-overlapping FVA ranges compared to baseline model	1207	1235
Of these, likely to have a significant effect on metabolism…	74	37
• **… due to differing by a factor of 1.1 or more**	72	37
• **… due to only carrying flux in one model**	25	5
• **… due to fluxes running in different directions**	0	0

3.5. Exploration of alternative engineering strategies using sampling of the flux solution space

FBA computes a single steady-state solution to a model which satisfies the objective. It can be useful, however, to capture all alternative solutions and their associated probability, also known as the solution space. This can help the identification of alternative pathways which may also be viable targets for PHB production. Flux sampling provides a way of looking at the entire solution space of the metabolic network under study by generating a sequence of feasible solutions (a 'chain') which satisfy the constraints until the entire sequence space has been analysed.

Convergence analysis is used to check that a chain is an accurate representation of the solution space, which requires the chain to contain enough samples. The output of flux sampling is a probability distributions of steady state reaction fluxes for each reaction in the model, thereby providing information both on the range of feasible flux solutions (similar to FVA) but also on their probability (*127*). Statistical tests can then be used to assess whether flux samples generated from two different models stem from the same distribution. Several flux sampling algorithms exist and have been extensively reviewed elsewhere (*127*). In this work the optimized general parallel (OPTGP) algorithm (*161*) was used because it is freely available in Python, converges fast and has a short run time (*127*).

As will be explained in section 3.7.1, cytosolic PHB engineering was selected as the target for this project and so flux sampling was only undertaken for cytosolic PHB. To look for alternative cytosolic PHB-producing pathways I set up the baseline and the cytosolic PHB-accumulating models for flux sampling. In both models the biomass accumulation reaction was constrained to the flux achieved in previous pFBA simulations where biomass accumulation was the objective. In those models, maximum biomass accumulation was limited by the constraint imposed on carbon fixation (Table S2.1). No such constraint was imposed during flux sampling, hence the need to constrain the biomass reaction directly. In the cytosolic PHB-accumulating model, PHB production was constrained to 10 % of biomass as previously described. The OPTGP algorithm was then used to generate sets of 5,000 samples using different thinning factors. Three convergence diagnostics were computed, and a high thinning constant was required to reach convergence, indicating the probable irregular solution

space of the model (*127*) (Table S3.1).

The flux distributions for each reaction across the solution spaces of the baseline and the cytosolic PHB-producing models were compared. The flux samples of a reaction were deemed to stem from different distributions if they satisfied a set of predefined conditions. Flux paths involving these reactions were analysed using a constraint scan and correlation analysis to identify reactions whose fluxes were significantly correlated over the range tested. Significantly correlated reactions were taken as forming part of the same metabolic pathway (for more details see section 2.1.6).

3.5.1. Flux sampling reveals expected relationships between reactions

Certain correlations between reactions confirm the flux paths already identified for cytosolic PHB accumulation using FBA. For instance, flux distributions of nocturnal vacuolar citrate storage were shown to be significantly different between the baseline and the cytosolic PHB-producing models (Figure 3.4). A constraint scan and correlation analysis of flux through the vacuolar citrate storage reaction showed that it was perfectly negatively correlated (r = 1.0) with flux through plastidial threonine synthesis, mitochondrial acetaldehyde import, peroxisomal acetate import, and peroxisomal citrate export; these are all reactions shown to contribute to peroxisomal citrate production (Figure 3.3). Thus, a lower amount of citrate stored in the vacuole results in a larger contribution of peroxisomal citrate towards cytosolic PHB production. This further supports the observation that boosting vacuolar citrate storage is enough for maximal PHB production without citrate export from the

peroxisome.

In addition, analysis of flux sampling solutions can reveal those reactions directly correlated with PHB synthesis. As expected, flux distributions of daytime PHB production were shown to be significantly different between the baseline and the cytosolic PHB-producing models (Figure S3.2). A constraint scan and correlation analysis of flux through the daytime PHB accumulation reaction revealed a perfect positive correlation (r = 1.0) with daytime ACL as well as the non-phosphorylating $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN). This reaction catalyses the oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate using the reduction of $NADP^+$ to NADPH, and therefore represents a variant of glycolysis that conserves energy as NADPH rather than as ATP. The correlation between cytosolic PHB production and GAPN activity is presumably due to the requirement of acetoacetyl-CoA reductase for NADPH (Figure 1.1A).

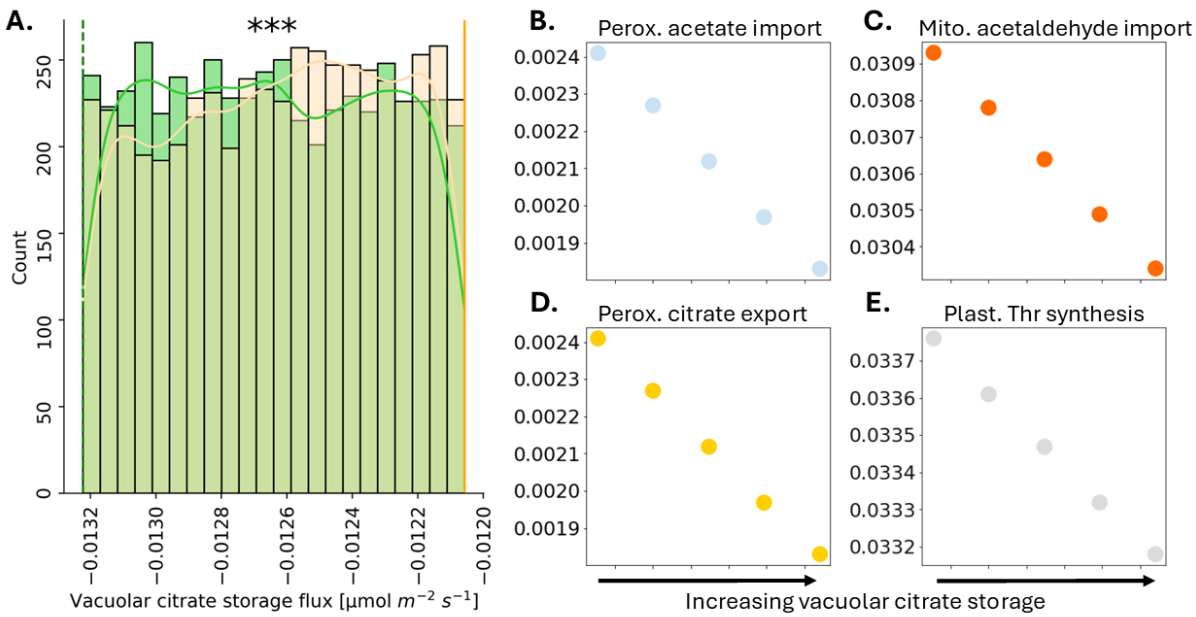


Figure 3.4: Flux sampling and correlation analysis of the vacuolar citrate storage reaction. (A) The flux distributions of the vacuolar citrate storage reaction in the baseline (orange) and cytosolic PHB-producing (green) models. Result of Kruskal-Wallis analysis of difference between distributions shown at the top of the figure. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05. Green dotted and orange lines represent the value carried by the reaction following pFBA in the cytosolic PHB and baseline models, respectively. Note that flux values are negative to denote reaction directionality, and that possible fluxes are limited by the constraint imposed on the reaction as explained in Table S2.1. (B-E) Results of correlation analysis between vacuolar citrate storage reaction and peroxisomal acetate import (B), mitochondrial acetaldehyde import (C), peroxisomal citrate export (D), and plastidial threonine synthesis (E). All fluxes are in units of µmol m^{-2} s^{-1}.

3.5.2. Alternative pathways predicted by flux sampling

Because the flux sampling algorithm does not use minimization of total fluxes as a secondary optimization objective (i.e. pFBA), solutions previously ignored due to a higher sum of fluxes, such as nocturnal PHB production, can emerge. Correlation analysis revealed that the key reaction to allow nocturnal PHB production was ACL, such that any nocturnal ACL activity in the range tested during the correlation analysis

resulted in nocturnal PHB production (Figure S3.3). The flux path to nocturnal PHB was analysed by constraining the flux through nocturnal ACL to the highest density point of its flux distribution (i.e. the most common value for flux through this reaction in the solution space) in the cytosolic PHB model and running pFBA (Figure 3.5A). As expected, this resulted in a slightly higher sum of fluxes (+ 0.3 %) than in the cytosolic PHB model without this constraint.

Nocturnal cytosolic PHB production was not limited by the constraint on vacuolar citrate storage and so did not require any peroxisomal citrate and relied solely on citrate exported from the mitochondrion. In addition, a nocturnal cytosolic PHB production scheme cannot directly use the products of photosynthesis and relies on carbon being stored during the day, particularly in the form of starch and sucrose. In the model starch storage was constrained and nocturnal carbon was mostly obtained from stored sucrose (Figure 3.2A). In the nocturnal PHB production scenarios, sucrose storage increased by 0.8 %; sucrose transport out of the vacuole is energetically costly, and hence this may partly account for the less efficient production of PHB at night.

Flux sampling also revealed a flux mode involving both daytime and nocturnal PHB production. This behaviour emerged as a feature of flux through peroxisomal malate synthase in the daytime. Peroxisomal malate synthase carried no flux in either the baseline or cytosolic PHB models already described. Constraint scanning and correlation analysis showed that flux through malate synthase above a certain critical value resulted in the simultaneous production of PHB in the day and in the night (Figure S3.3). The flux through malate synthase was constrained to this value and flux paths identified after pFBA as previously described. The proportions of total PHB accounted

for by day and night production schemes were 70.2 % and 29.8 % respectively. The flux paths to night and daytime cytosolic PHB were very similar to the isolated day and night production scenarios. During the night, PHB production proceeded as in the case of flux through nocturnal ACL, albeit with lower flux through ACL and no production of acetylserine from acetyl-CoA. During the day, the only notable but significant difference was the lack of citrate export from the peroxisome (Figure 3.5B). Nocturnal PHB production seems able to compensate for daytime peroxisomal citrate export, albeit less efficiently. Indeed, in this model configuration the sum of fluxes increased by 4.8 % compared to the cytosolic PHB model without this constraint.

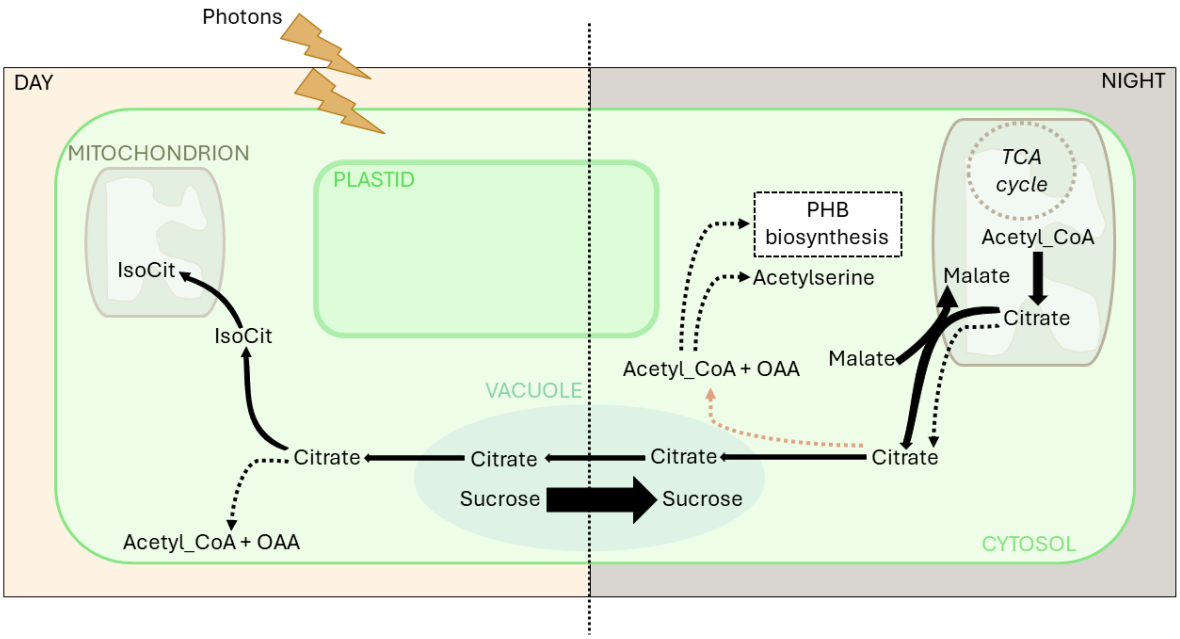
Flux through peroxisomal malate synthase at or above this critical value led to a day and nocturnal PHB flux mode because it precluded peroxisomal citrate synthase activity and thereby peroxisomal citrate export, which usually contributed to the total yield of daytime PHB. This is because the model is unable to produce enough acetyl-CoA in the peroxisome to satisfy both the constraint on malate synthase and flux through citrate synthase. Instead, the synthesis of PHB at night is used to meet the constraint on PHB flux, given the constraint on the vacuolar citrate storage. Indeed, removing the constraint on citrate storage removes the production of PHB at night. This illustrates again how important the experimentally derived citrate constraint is on the behaviour of the model, and how important increasing citrate storage in the vacuole is for ensuring an efficient path to PHB production.

3.5.3. Flux sampling conclusion

FBA computes single steady-state solutions which satisfy the objective, and the range

of these solutions can be analysed using FVA. This approach provides no indication as to which solutions within the range are most likely, however, or whether all solutions within the range computed by FVA are feasible. Flux sampling can explore the feasible steady state flux space in a metabolic network. As such reactions shown to carry flux using flux sampling can be considered both feasible and characteristic of the metabolic network under study (i.e. the reaction does not just carry flux in the one steady-state solution computed by FBA). Flux paths previously identified as important for the cytosolic accumulation of PHB using pFBA (such as vacuolar citrate storage and daytime ACL) were also identified as important for cytosolic PHB production using flux sampling (Figures 3.4 and S3.2). In addition, flux sampling revealed that, although less efficient, several different nocturnal PHB production modes are feasible. This is an important finding given that the constitutive expression of the PHB-producing enzymes is likely to occur at night too. To conclude, the conclusions from flux sampling allows us to be more confident in the key behaviours of the model when set up to produce cytosolic PHB.

A.



B.

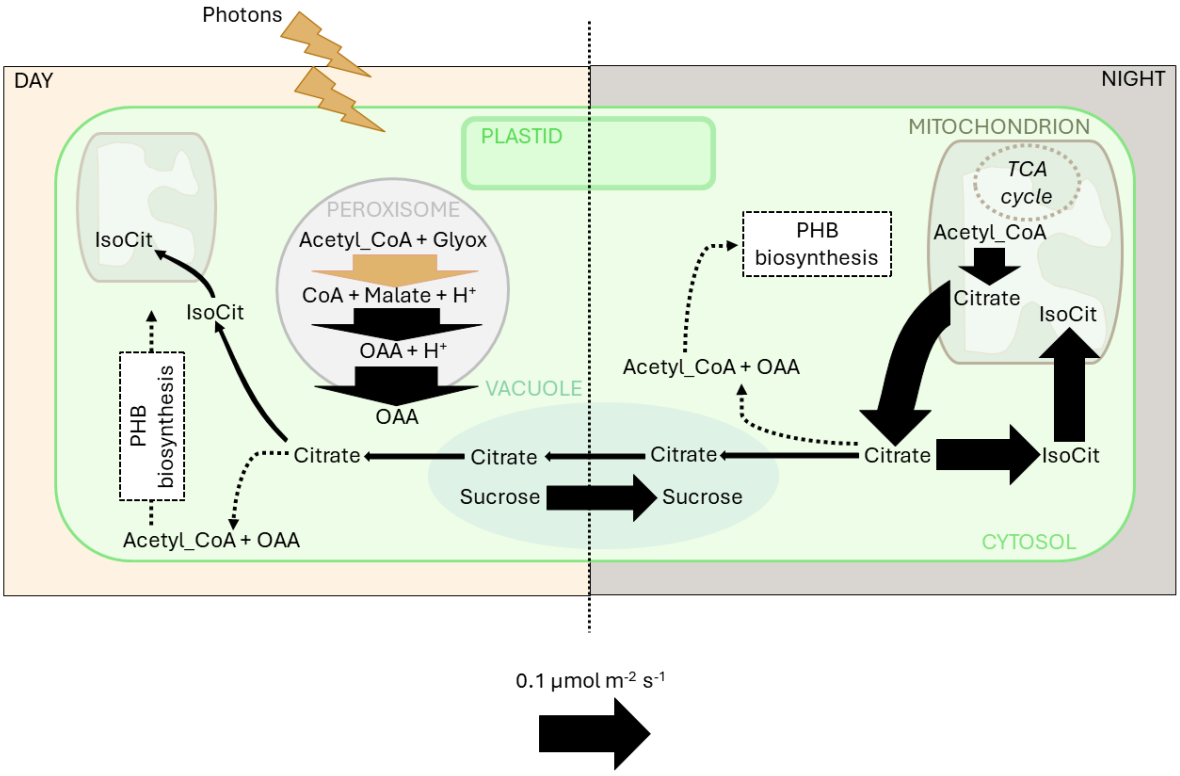


Figure 3.5: An illustration of main flux paths to PHB synthesis when the model is constrained based on flux sampling to carry flux through (A) nocturnal ACL, or (B) daytime peroxisomal malate synthase. Constrained fluxes are shown as orange arrows. Width of black arrows are scaled according to the flux they carry in units µmol m^{-2} s^{-1}. Fluxes below 0.01 are represented using a dotted line. α-KG, α-ketoglutarate; OAA, oxaloacetate; IsoCit, isocitrate; Glyox, glyoxylate.

3.6. Investigation into the effect of sucrose on model conclusions

Although *Marchantia* is capable of fully autotrophic growth, its rate of photosynthesis is low and, in many cases, supplemental sucrose was provided in the growth medium of plants for the experimental studies. This was to accelerate growth and to facilitate the production of gemmae in reasonable timescales. For this reason, the constraint on sucrose import (Table S2.1) was relaxed and the impact on flux paths to cytosolic PHB accumulation investigated. Fully relaxing the constraint on sucrose import during the day and the night resulted in unrealistic flux patterns; being the more reduced carbon source, sucrose was preferentially imported resulting in no photosynthetic assimilation of CO_2. As a more realistic starting point a constraint was imposed on the model such that the sum of daytime and nocturnal sucrose uptake flux was half of CO_2 daytime uptake rate. The value for sucrose import obtained using this constraint was used as the starting point for a constraint scan. Briefly, the sum of night and daytime sucrose uptake was constrained to this value and the fluxes to key reactions after pFBA were monitored. The constrained value was decreased by 10 % each iteration until it reached 0. The results of the scan are shown in Figure 3.6.

At low values of total sucrose import, sucrose was preferentially imported during the night. As shown in Figure 3.3A, the cytosolic-PHB model uses sucrose as a source of carbon during the night due to highly constrained daytime starch storage. Only at relatively high sucrose import values did the model import sucrose during the day (Figure 3.6A). Storing sucrose in the vacuole carries a cost given that it requires protons

to be pumped into the vacuole using ATP which may explain the preference for nocturnal import. However, the model has a limited ability for absorbing carbon at night given constraints on biomass accumulation and citrate storage and must switch to import during the day at high total sucrose import values.

The change from night-only to night and daytime sucrose import also corresponded to the point at which some ACL activity and thereby PHB production also happened in the night as opposed to solely during the day. As with the malate synthase situation previously described (section 3.5.2), the flux of citrate from nocturnal ACL took the place of daytime citrate export from the peroxisome, such that these two traces are mirror images of each other (Figure 3.6B). Presumably the model uses the production of PHB at night (which effectively acts as a carbon sink) as another way to absorb the excess carbon as a result of the sucrose uptake constraint.

In general, allowing sucrose import did not change the main conclusions from the model and cytosolic PHB was produced during the day via ACL from citrate mostly exported from the vacuole (as shown in Figure 3.3A). For this reason, the possible engineering targets identified from the model with sucrose import constrained were assumed to still hold for plants grown on media containing sucrose.

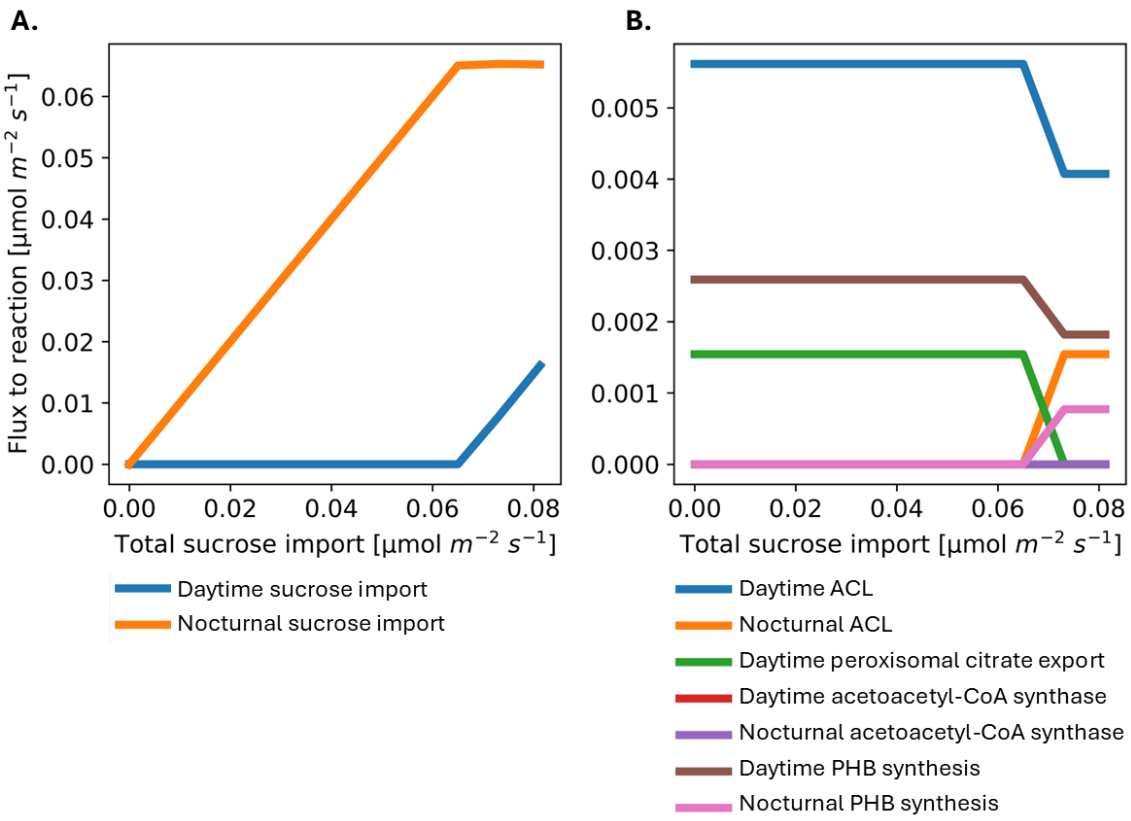


Figure 3.6: The effect of increasing total sucrose import on selected fluxes, including on daytime and nocturnal sucrose import (A), and reactions previously shown to be involved in cytosolic PHB synthesis (B).

3.7. Discussion

3.7.1. Model analysis suggests that engineering the biosynthesis of PHB in the cytosol is a more promising strategy than plastidial production

In this work, pFBA was used to find the optimal predicted paths to the production of PHB in the cytosol and in the plastid of *M. polymorpha* (Figure 3.3). To address the limitation that fluxes in an FBA solution are not necessarily uniquely defined, FVA was

used to identify all fluxes in the PHB-producing models that differed from the baseline model. When combined with other considerations, such as experimental limitations or results, this information can be used to determine the most promising subcellular location for the engineering of PHB in *M. polymorpha*.

Using FVA, it was determined that slightly fewer reactions compared to baseline *M. polymorpha* metabolism needed to change for cytosolic production than for plastidial production (Table 3.2). However, the reactions in the cytosolic PHB model needed to change more and in more significant ways than the reactions predicted to differ between the baseline and plastidial-PHB production models.

An additional consideration, however, is that in the case of plastidial production the model predicted fluxes through several subcellular compartments and a route that involved highly toxic acetaldehyde as well as threonine which is present in low concentrations in the cell. These attributes do not lend themselves well to engineering. In contrast, in the case of cytosolic production, the majority of the citrate required for acetyl-CoA synthesis (88.3 %) was provided via nocturnally stored vacuolar citrate, mimicking a pattern of carboxylic acid accumulation that is known to occur already and thereby providing a pathway much more amenable to engineering. In addition, if the constraint on the upper bound of vacuolar citrate storage was removed all citrate used for PHB synthesis was obtained from the vacuole, showing that boosting the vacuolar citrate storage capacity by 11.7 % was sufficient for maximal citrate production under these conditions. The biggest change in flux associated with PHB production using this pathway was flux through ACL, providing a potential single target enzyme able to have a large impact on the ability of the network to support cytosolic

PHB production. Finally, cytosolic PHB production avoids the potential interference with photosynthesis due to light scattering identified in plants accumulating PHB in the plastid and does not directly compete with fatty acid biosynthesis in the chloroplast for acetyl-CoA. On balance, cytosolic PHB engineering was determined to be more favourable and was taken forward as the target location in this project.

3.7.2. Possible engineering targets for engineering cytosolic PHB production in *M. polymorpha*

Engineering targets for PHB production can be inferred from the model of baseline *M. polymorpha* metabolism, the flux paths towards PHB production, and the results of flux sampling, and are summarized in Table 3.4. In this first case, flux though central metabolism can be compared before and after experimentally derived constraints on organic acid transport are applied. In the absence of a constraint on the vacuolar storage of fumarate and malate, these reactions carried significant flux in the baseline model. The flux mode involving malate and fumarate storage was more efficient, which was underlined by a lower sum of fluxes, although it led to a small (- 1.4 %) decrease in flux to biomass (Table 3.1). The reduced sum of fluxes was partly accounted for by simplified citrate metabolism; all citrate exported from the mitochondrion in the night was translocated to the vacuole leading to a lower overall citrate efflux from the mitochondrion, and a malate-citrate antiporter was used to export citrate from the vacuole during the daytime (Figure 3.2C).

Given that malate and fumarate storage in the vacuole results in a more efficient flux path through central metabolism, the vacuolar storage of malate and fumarate could

be an engineering target. Vacuolar transport of malate and fumarate have been relatively well studied. The transport of malate across the vacuolar membrane in *A. thaliana* is mediated by a tonoplast malate/citrate antiporter (*At*tDT) and two members of the aluminium activated malate transporter (ALMT) family, *At*ALMT9 and *At*ALMT6 (*205*). The cytosolic fumarase FUM2 from *A. thaliana* has been shown to be essential for the accumulation of fumarate in *A. thaliana* leaves, and a positive linear correlation was shown between fumarase activity and accumulated fumarate levels indicating that fumarate was being sequestered in the vacuole (*206*). These transporters could be targets for overexpression to encourage the storage of malate and fumarate in the vacuole of *M. polymorpha*.

By comparing the baseline model to the cytosolic PHB-accumulating model I can identify where differences in the metabolic network exist and thereby where engineering interventions are needed. PHB was synthesized in the cytosol during the day from acetyl-CoA obtained primarily from nocturnally stored vacuolar citrate (Figure 3.3A). In addition, the proportion of citrate not derived from the vacuole (11.7 %, from peroxisomal export) was a result of the constraint imposed on citrate storage, such that removing this constraint resulted in all cytosolic citrate used by ACL being exported from the vacuole. Given that this pathway mimics a known diurnal pattern of carboxylic acid accumulation, increasing citrate storage and/or release is a promising engineering target. Unfortunately, knowledge about plant vacuolar citrate transporters is limited. The *At*tDT malate/citrate antiporter from *A. thaliana* has been shown to transport citrate and malate in a 1:1 antiport mechanism; ion preference indicates it might function predominantly to export citrate as it preferentially binds the dianion

form, which is the predominant form inside the vacuole (*181*). However, it is not thought to be the main tonoplast citrate carrier in *Arabidopsis* such that at least one other plant tonoplast citrate carrier remains unidentified (*207*). In general, studies on the regulation of citrate transport are very limited (*208*).

A complementary strategy to encouraging citrate export from the vacuole would be to increase the nocturnal vacuolar storage of citrate. Encouraging the transport of citrate into the vacuole might sequester it from catabolic reactions and allow the export of increased levels of citrate during the day when PHB synthesis is more favourable. One approach to achieve this would be to take advantage of the fact that a lower vacuolar pH has been shown to favour citrate import into the vacuole in a phenomenon known as the 'acid trap' (*205*). For this a proton pump would need to be overexpressed – however, the V-type ATPases that function in plants are very large and multi-subunit. P-type ATPases located on the plasma membrane are formed of a single subunit, and have also been shown to be localized to the tonoplast in *Petunia hybrida* (PH5/PH1 pair) *(209),* apple (*Ma*10) (*210*), and *A. thaliana* (*At*AHA10) (*211*). It is likely, however, that acidifying the vacuole would lead to the disruption of vacuolar accumulation of other metabolites; the result of this on metabolism might have to be considered. For instance, malate transport through the *At*ALMTs occurs only in the presence of an electric potential gradient across the tonoplast; however, this is expected to decrease with a decrease in vacuolar pH (*205*).

Finally, the citrate stored in the vacuole at night was exported from the mitochondrion; hence another way to increase citrate flux towards PHB might be to encourage this export reaction. DIC2 from *A. thaliana* was recently shown to be a mitochondrial

malate importer/citrate exporter (*212*) and could be an overexpression target.

FVA and comparative analysis of baseline and cytosolic PHB-producing models revealed that the largest difference in flux of any reaction was the 12.7-fold increase through ACL with PHB synthesis. The overexpression of ACL thus presents itself as an important target for ensuring PHB production. In the cytosol, ACL catalyses the ATP and CoA-dependent cleavage of citrate to form acetyl-CoA and oxaloacetate. In *A. thaliana,* ACL is a heterooctamer formed of the two subunits ACLA and ACLB, encoded by three and two genes, respectively (*111*). However, human ACL (ACLY) has been shown to exist as a homotetramer (*213*) and so could be a more promising target for increasing cytosolic acetyl-CoA provision.

All suggested engineering targets are shown in Table 3.4. The engineering strategy tested *in planta* will be discussed in section 5.3.

Table 3.4: Possible engineering targets for the cytosolic production of PHB in *M. polymorpha*.

Engineering goal	Possible target
Increase ACL activity	Overexpression of homomeric ACLY from *H. sapiens*.
Increase vacuolar citrate storage	- Acidification of vacuole Expression of vacuolar P-type ATPases, for instance PH5 from *P. hybrida*
Increase vacuolar citrate export	Expression of the *At*tDT malate/citrate tonoplast antiporter from *A. thaliana*
Increase mitochondrial citrate export	Expression of the DIC2 malate/citrate mitochondrial antiporter from *A. thaliana*
Increase vacuolar malate and fumarate storage	- Vacuolar malate import Expression of vacuolar malate transporter, such as *At*ALMT9 from *A. thaliana*. - Vacuolar fumarate import Expression of cytosolic fumarase FUM2 from *A. thaliana*

3.7.3. Previous attempts at engineering acetyl-CoA availability and PHB synthesis

The nocturnal storage of vacuolar citrate and the subsequent conversion of this citrate to acetyl-CoA was the key differentiator between a biomass-only and a PHB-biosynthesis model configuration (Figures 3.3A and 3.4). Given that acetyl-CoA is a key metabolic intermediate and a precursor for many groups of high-value compounds (terpenoids, fatty acids, flavonoids etc), increased acetyl-CoA provision has been the target of many metabolic engineering approaches (*214*). These approaches range from the design of novel pathways to the overexpression of enzymes catalysing reactions already present in the host. For example, a pathway based on a phosphoketolase from *Bifidobacterium breve* and a phosphotransacetylase from *Clostridium kluyveri* was implement in the cytosol of *S. cerevisiae*. Phosphoketolase can split fructose-6-phosphate and xylulose-5-phosphate into acetyl-phosphate, thus redirecting part of the OPPP towards generation of acetyl-CoA. Using this approach free fatty acid titre was improved by approximately 30 % (*215*).

The engineering of *S. cerevisiae* for heterologous production has also made use of the cytosolic overexpression of ACL to modulate acetyl-CoA provision. For instance, the overexpression of ACL from *A. thaliana* increased the production of 1-hexadecanol by 55 % (*216*). Similarly the overexpression of ACL from *Yarrowia lipolytica* was able to increase yield of 1-hexadecanol by 136 % (*216*), cause a 2-fold increase in *n*-butanol productivity (*217*), and increased fatty acid productivity by 26 % (*218*). The overexpression of ACL has also been combined with efforts to increase citrate provision. These attempts have all involved manipulation of endogenous isocitrate

dehydrogenase (IDH) activity; IDH catalyses the conversion of isocitrate to α-ketoglutarate and is known to control rates of citrate catabolism (*219*). Also in yeast, disruption of both *idh1* (cytosolic) and *idh2* (mitochondrial) genes and cytosolic overexpression of ACL from *Mus musculus* resulted in a 5-fold increase in intracellular citrate, as well as a 92 % improvement in 9-hexadecanoic acid productivity (*219*). This approach was also shown to directly affect acetyl-CoA levels. CRISPR-Cas9 was used to disrupt *idh1* and *idh2* both separately and together, and cytosolic overexpression of ACL from *Y. lipolytica* resulted in a 2.5-fold increase in acetyl-CoA levels as well as a 37.1 % improvement in fatty acid productivity. Interestingly, even though mitochondrial citrate export wasn't manipulated, the *idh1* and *idh2* single mutants had the same intracellular and extracellular citrate contents as well as the same increase in acetyl-CoA contents over WT (*220*). This shows that a decrease in mitochondrial IDH activity is sufficient for higher cytosolic citrate availability to ACL.

Availability of acetyl-CoA has also been a target for the engineering of higher PHB yields. A similar method to the phosphoketolase/phosphotransacetylase approach already described has been used to increase the yield of PHB in *S. cerevisiae*. The reconstruction of the phosphoketolase pathway from *Aspergillus nidulans* in the cytosol resulted in a 6.9-fold increase in yield of PHB. However these strains showed a decrease in specific growth rate and biomass yield (*193*). In addition, analysis of a stoichiometric model of *S. cerevisiae* expressing the three enzymes required for PHB synthesis suggested that the theoretical yield of PHB could be improved by including a combination of ACL and a transhydrogenase reaction to catalyse the conversion of NADH and $NADP^+$ to NAD^+ and NADPH, respectively (*194*). Finally, ACL has been

overexpressed in *A. thaliana* engineered to produce PHB and was shown to rescue the negative phenotype associated with PHB accumulation in this plant. In this system a 2.7-fold increase in ACL activity over WT was sufficient for the ablation of a PHB-associated phenotype (*61*).

The overexpression of ACL either alone or in combination with other targets has thus been used extensively in metabolic engineering to increase provision of acetyl-CoA to target compounds, including PHB. The analysis of the model of *M. polymorpha* undertaken here shows that this is a promising target for increasing PHB yield in this species too.

The remainder of the targets identified for enhanced cytosolic PHB production are transporters (Table 3.4). Unfortunately, knowledge regarding plant transporters remains somewhat limited. Consider, for example, that of the 1500 published unique membrane protein structures only 51 are unique membrane protein structures from terrestrial plants and only 18 are plant transporters (as of November 2022) (*221*). While a structure is not essential for determining reaction mechanism, the lack of plant transporter structures is symptomatic of a lack of knowledge regarding plant intracellular transporters. Attributes such as substrate specificity, directionality, stoichiometry, and co-transport mechanism can all substantially influence flux through a metabolic model. Given that knowledge regarding plant transporters is limited, very few metabolic engineering attempts make use of the overexpression of plant transporters.

Finally, it is worth considering ways other than the direct manipulation of acetyl-CoA provision for increased flux of carbon into PHB production. A gene from

Synechococcus encoding a bifunctional enzyme with fructose-1,6-bisphosphatase (FBPase) and sedoheptulose-1,7-bisphosphatase (SBPase) activities was introduced into switchgrass lines producing plastidial PHB. FBPase and SBPase function in the CBB cycle and affect both sugar partitioning and carbon fixation, and the expression of genes encoding these proteins in tobacco has been shown to increase photosynthesis and plant growth (*222*). The levels of PHB as a % per unit dry weight after 2 months growth in the greenhouse in switchgrass plants expressing FBPase and SBPase were increased by 2.2- and 2.1-fold in mature and developing leaves, respectively (*47, 58*). This shows that increasing the efficiency of biomass accumulation can ultimately improve PHB yield.

3.7.4. Limitations of the modelling approach

The simplicity of stoichiometric models facilitates their construction as well as reduces computational load and thereby allows much bigger networks to be analysed. However, this approach is unable to model more complicated effects on metabolism such as changes in enzyme levels or metabolite pools as well as regulatory effects such as those caused by gene-regulatory responses to metabolic engineering interventions. For instance, the activity of ACL *in planta* is thought to be regulated post-translationally by sink strength (*61*). As such the production of PHB itself might be sufficient for the increased flux through ACL predicted in this model and the overexpression of ACL may have a limited effect on final PHB yield. Indeed, the overexpression of ACL in PHB-producing *Arabidopsis* lines rescued growth phenotypes but had no effect on yield of PHB expressed per unit dry weight (*61*). Similarly, as explained, previous stoichiometric models have predicted ACL to be a

good overexpression target in increasing target molecule accumulation. However, these benefits have not always been realized *in vivo*. For instance, the expression of an ACL from *A. thaliana* in yeast engineered for heterologous patchoulol synthesis was unable to increase product titre, which the author attributed to suboptimal kinetic properties and substrate affinity in the yeast host (*223*). In addition, intracellular citrate concentrations have been shown to have major effects on nuclear-encoded transcript abundance in *A. thaliana*, including on regulation of the TCA cycle, nitrogen and sulphur metabolism, as well as DNA synthesis (*224*). This reflects the fact that metabolites, as well as being involved in metabolic reactions, also have signalling functions, many of which are still unknown. These and other effects are not considered in the simplified model of metabolism constructed here, which can result in poor performance as a metabolic prediction tool.

In addition, the modelling approach here was limited to studying the effects of PHB production on known reactions in the central metabolism of *M. polymorpha*. In this way reactions already known to occur could be targets for overexpression. An alternative strategy could have involved the introduction of heterologous pathways known to, for example, increase acetyl-CoA provision *in vivo*. Examples already mentioned include the phosphoketolase/ phosphotransacetylase approach for redirecting OPPP flux towards acetyl-CoA (*193*), a transhydrogenase reaction for reducing-power provision (*194*), or even the deletion of IDH activity (*220*). The model could have been modified according to these findings and effects on metabolic fluxes as well as yield investigated.

Finally, the accuracy of model predictions is fundamentally dependent on the quality

of the model and its constraints. The behaviour of the model both for biomass production and in the case of cytosolic PHB accumulation were highly dependent on the constraints imposed on organic acid storage. In the baseline model, allowing the vacuolar storage of malate and fumarate resulted in significant storage of nocturnal malate and daytime fumarate. It led to more efficient flux through central metabolism and resulted in simpler citrate storage behaviour (Figure 3.2C and Table 3.1). Experimental data regarding patterns of organic acid storage in *M. polymorpha* clearly reveals an absence of storage of malate and fumarate (see https://github.com/AnyaLindstromBattle/Thesis). This is unusual and most plants are known to store one or both of these organic acids (*200, 225*). In addition, previous work has suggested that CAM cycling is required for a model of *Marchantia* to achieve growth rates in line with experimentally determined values. CAM cycling relies on the storage of nocturnal malate in the vacuole which is exported during the day and converted to pyruvate and other compounds to be re-assimilated in the CBB cycle (*140*). Similarly, the nocturnal storage of citrate in the vacuole was introduced as a constraint in the model based on experimental data (see https://github.com/AnyaLindstromBattle/Thesis), and the presence of this constraint also had significant effects on the fluxes through the model. Clearly more work needs to be done to unravel the storage of organic acids in *M. polymorpha*, especially given the effect these constraints have on model outcomes.

Nevertheless, the work presented in this chapter represents a first attempt at investigating PHB synthesis in *M. polymorpha*. Analysis of a model representing the central metabolism of *M. polymorpha* has revealed possible engineering targets both

novel and in line with previous attempts at engineering PHB synthesis or similar metabolism. The inherent limitations to the predictive power of stoichiometric models further highlights the need for a fast testbed species in which model-driven metabolic engineering targets can be tested.

4. Development of *Marchantia polymorpha* as a metabolic engineering platform

4.1. Introduction

The recent development of tools and techniques for the genetic engineering of *M. polymorpha* (*34, 35, 38, 40, 42, 226, 227*) has promoted its emergence as a possible testbed species for metabolic engineering. However, knowledge regarding the manipulation of this species is still limited when compared to established model plant species such as *A. thaliana*. As a result, certain protocols such as for transformation have not been optimized so as to be reproducible in different contexts. In addition, genetic elements such as promoters remain poorly characterized in *M. polymorpha*. To date only the p*Mp*EF1α and the p35S promoters have been extensively tested for constitutive expression in this species (*228*).

Inducible promoters have been used in plant metabolic engineering to overcome deleterious or lethal effects of product accumulation and metabolic disruption. Given the negative phenotypic effects associated with PHB production in plants, thought to be caused primarily by the competition between PHB and biomass accumulation for acetyl-CoA, I hypothesized that inducible expression might help overcome negative effects on plant growth by temporally separating PHB and biomass production.

Inducible promoters tested in plants include those activated by small molecules; of these one of the most widely used is the ethanol-inducible *AlcA* promoter (*67*). In *Marchantia*, an oestrogen-activated (*229, 230*) and an auxin-responsive (*231*) promoter have been tested.

However, chemically induced systems suffer from a range of drawbacks. For example, they depend on the inducer diffusing into tissues, organs, and cells, which limits spatial and temporal resolution both of activation and of inactivation of expression. They can also be expensive and interfere with normal cellular activities due to chemical crosstalk or toxicity (*232*). The control over gene expression using light, known as optogenetics, does not suffer from these drawbacks and allows for the precise spatiotemporal control over gene expression with low toxicity. Optogenetics has reached widespread use in fields such as mammalian cell biology, but in plant science its usage has been limited. Most of the optogenetic tools adapted or developed for plants are incompatible with growth of plants in white light without promoter activation. This is because they are typically activated by a specific wavelength of light forming part of the white light spectrum needed for normal plant growth (*233*). In contrast, the PULSE system enables growth of plants without promoter activation in normal white light conditions; this is because it combines a red-light inducible module with a repressor module that prevents gene expression in all light conditions apart from monochromatic red light. This system has been shown to be fully functional in *Arabidopsis* protoplasts, transiently infiltrated *N. benthamiana* leaves, and in several stably transformed *Arabidopsis* lines (*73*).

In this chapter tools and techniques contributing to the use of *M. polymorpha* as a

testbed species in metabolic engineering are described. Adaptations to existing protocols as well as key information regarding the transformation and growth of *M. polymorpha* are explored. In addition, this chapter describes the adaptation and application of the PULSE optogenetic system in *M. polymorpha*. The suitability of PULSE for inducible PHB biosynthesis is also explored. The work in this chapter represents a significant expansion of the toolbox for the metabolic engineering of *M. polymorpha*.

4.2. Using *M. polymorpha* as a testbed species for metabolic engineering

4.2.1. Transformation of *M. polymorpha*

Protocols currently exist for the *Agrobacterium*-mediated transformation of the thallus, gemma, or spore stages of the *M. polymorpha* lifecycle *(34, 35, 227)*. To gain an insight into the adoption of these techniques by the *Marchantia* community, a comprehensive literature review was undertaken. Briefly, the search terms 'transformation', 'Marchantia', and 'polymorpha' were used in a Web of Science database search (Clarivate™ [Web of Science™]. © Clarivate 2024. All rights reserved.). This search returned 252 results (as of March 2024) and the primary research articles were surveyed for the type of transformation method used (Figure S4.1).

Analysis of the literature revealed that very few examples of successful transformation of gemmae have been published. Indeed only 11 publications exist in which successful transformation of gemmae is reported, and of these five publications are

reporting the method itself or improvements to this method. The transformation of gemmae was unsuccessful in my hands and generated no transformants, and a general consensus has emerged that the transformation of gemmae is very difficult (Susanna Streubel, formerly at the Gregor Mendel Institute, Vienna, Austria [personal communication]).

The transformation of thalli and of spores of *M. polymorpha* have been equally successful, with 71 and 75 publications reporting transformation using these two techniques, respectively (Figure S4.1). This suggests that the transformation of thalli and spores should be the focus of future studies on *Marchantia*. In addition, these two methods are favoured for different applications. Whereas sporeling transformation is advantageous when wanting to generate many lines for screening, most publications using thallus transformation were super-transformation or complementation studies, or other applications where fewer lines were needed. Indeed, thallus transformation is labour intensive and typically around 100 thallus fragments are required for one transformation.

Thallus and sporeling transformation methods were used in this project. The transformation of thalli relies on the induction of regeneration by removal of the apical part of the thallus. The large amount of cell division predicted to occur in the regenerating thalli is conducive to *Agrobacterium* infection (*234*). In my hands thallus transformation was successful albeit at lower efficiency than published (15-20 % compared to 63 – 72 % in the published literature) for WT Tak1 plants. No transformants were obtained for WT Tak2 plants, reflecting an understanding that Tak1 is easier to transform using this method (Susanna Streubel, formerly at the Gregor

Mendel Institute, Vienna, Austria [personal communication]). For the thallus transformation of mutant lines, growth steps had to be modified to match the slower growth rates of the mutants compared to WT plants. Specifically, thalli were grown for 3 weeks instead of 2 weeks before thallus cutting, and the basal fragments were allowed to regenerate for 5 days rather than 3 days. In addition, there was a link between transformation efficiency and plant health, such that the line with the highest production levels of PHB resisted super-transformation with this method (see Chapter 5). With two lines producing PHB at moderate levels, the transformation efficiency was still reduced compared to WT, at around 10 %.

The transformation of the spore stage of *M. polymorpha* was the most reliable and efficient at generating transformants. Using this method several hundred transformants were typically obtained for each transformation. There are hundreds of thousands of spores in a sporangium (*235*), and several sporangia are used for each transformation. The increased number of transformants achieved using this method is probably a reflection of the larger amount of starting material rather than an increase in transformation efficiency compared to the transformation of thalli.

4.2.2. *M. polymorpha* is insensitive to kanamycin

The kanamycin resistance encoded by the neomycin phosphotransferase II gene (*nptII*) of Transposon Tn5 is the most widely used antibiotic selection marker in dicotyledonous plants. In this work it was established that kanamycin is unsuitable for use as a selectable marker with *M. polymorpha*. Although gemmae freshly isolated from gemmae cups were killed by kanamycin concentrations at or above 50 μg mL^{-1},

when gemmae were grown in normal conditions for 4 days prior to selection (as in the gemmae transformation protocol (*227*)), up to 25 % of gemmae survived and produced green outgrowths. However, none of these plants were true transformants as was indicated by negative PCR genotyping results. The same was true when kanamycin was used as the selectable marker in thallus transformation, where no positive genotyping results were obtained for any apparent transformant either. This indicates that older *Marchantia* tissue has resistance to kanamycin. This has precedence in monocots which are known to be insensitive to kanamycin (*236*). This effect is known but not published (Susanna Streubel, formerly at the Gregor Mendel Institute, Vienna, Austria [personal communication]).

4.3. A fully functional optogenetic system for light-controllable transgene expression in *M. polymorpha*

4.3.1. The PULSE system

PULSE is an optogenetic system that is activated with monochromatic red light and inactive in ambient white light (*73*). Its activity relies on the combinatorial activity of two engineered photoreceptors (Figure 4.1). The first, B_{off}, represses gene expression under blue light. It is built from the EL222 photoreceptor from *Erythrobacter litoralis* which is composed of a light-oxygen-voltage (LOV) domain and a helix-turn-helix (HTH) domain, and which binds its target DNA sequence C120 as a dimer upon blue light illumination. In B_{off}, EL222 is fused to the transcriptional repressor domain SRDX, and binds to a reporter module under the control of a quintuple-repeat of C120 ($[C120]_5$) flanked by the enhancer sequence of p35S and the minimal domain of the human

cytomegalovirus constitutive promoter ($P_{hCMVmin}$). In the absence of blue light, the engineered EL222 is folded and unable to bind $(C120)_5$, resulting in expression through the promoter. In blue light EL222 binds $(C120)_5$ and prevents expression.

The second element of PULSE is the R_{on} module, which activates gene expression in the presence of red light. R_{on} is based on the red/far-red light-dependent interaction of phytochrome B (PhyB) and phytochrome-interacting factor 6 (PIF6) from *A. thaliana* (*237*). Amino acids 1-100 from PIF6 are fused to a DNA binding domain (E) that binds its operator site $(etr)_8$. Amino acids 1-650 from PhyB are fused to an activator domain (VP16) and a nuclear localisation sequence such that, in red light, the PhyB fusion protein is recruited to PIF6 bound to the promoter site to turn on gene expression. In the presence of far-red light PhyB undergoes a conformational change and is no longer able to bind PIF6; gene expression is thereby turned off.

PULSE combines B_{off} and R_{on} by the construction of a synthetic target promoter, P_{Opto}, which integrates the $(C120)_5$ and $(etr)_8$ binding domains upstream of $P_{hCMVmin}$ to drive expression of the target gene, such as firefly luciferase (FLuc). As a result of the combinatorial effect of the B_{off} and R_{on} switches, blue or white light will prevent gene expression (due to binding of the repressor B_{off}), far-red light will do the same (due to inactivity of the activator), but in red light R_{on} will bind $(etr)_8$ and activate gene expression.

As such, as well as carrying the synthetic P_{Opto} promoter, the PULSE system relies on the constitutive expression of the PhyB, PIF6, and EL222 modules, which are under the control of minimal p35S promoters. Finally, the system includes a luciferase reporter normalisation element constructed from *Renilla* luciferase under the control of the

constitutive polyubiquitin 10 gene promoter from *A. thaliana* (p*At*Ubi10). *Renilla* luciferase catalyses the oxidative decarboxylation of coelenterazine resulting in the emission of blue light (*238*).

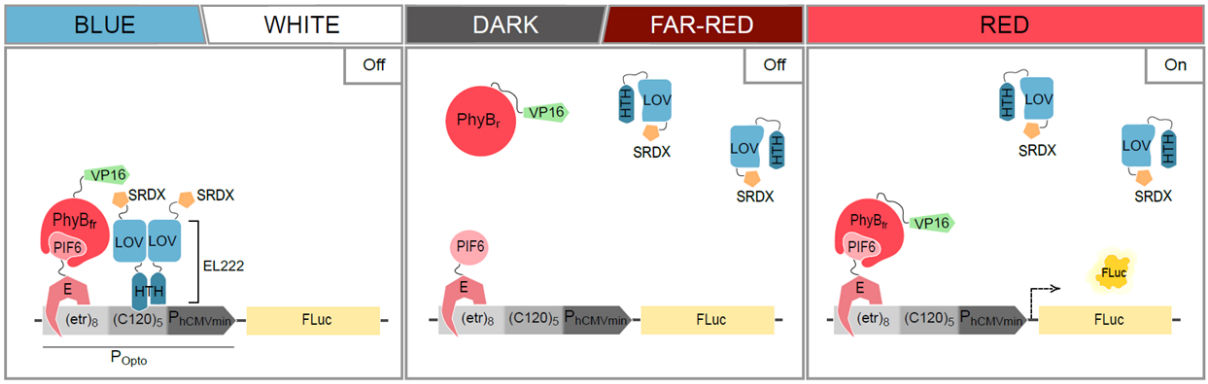


Figure 4.1: The PULSE optogenetic system and its activity under different light conditions. See text for abbreviations used. Reproduced with permission from Springer Nature (https://www.nature.com/nmeth/) from (*73*).

4.3.2. Characterisation of PULSE in *M. polymorpha*

The PULSE system was modified for use in *M. polymorpha* by the replacement of kanamycin resistance with a hygromycin resistance cassette and transformed into plants using sporeling transformation. All *E. coli* growth steps were carried out at 30 °C due to the presence of repetitive sequences and their propensity for mutation (Matias Zurbriggen, Heinrich Heine University, Düsseldorf, Germany [personal communication]). Transformants were screened by the exogenous application of D-luciferin and imaging of luminescence using a photon counting camera after 4 h in monochromatic red light, followed by genomic DNA extraction and genotyping for

confirmation of the integration of the PULSE system into the M. *polymorpha* genome.

4.3.2.1. Dynamics and distribution of PULSE in M. polymorpha

To confirm that the expression of luciferase was induced as a response to red light, four lines confirmed for the presence and expression of the PULSE construct (lines #18, 8, 27, 37) were grown from gemmae in clear 24-well plates for 15 d. After addition of exogenous D-luciferin, plants were placed in white light for 12 h followed by 12 h in red light. Finally, plants were subject to white light for another 24 h. Images and photon counts were captured every 2 h during this period by a photon counting camera and photon counts normalised by final plant mass.

There was a clear increase in luminescence, and thereby luciferase expression, as a result of the red-light treatment (Figure 4.2A). Normalised luminescence differed significantly between lines, and lines #27 and 37 showed higher levels upon induction than lines #8 and 18. WT Tak1 and Tak2 plants displayed negligible levels of normalised luminescence throughout the course of the experiment. The level of activity through P_{Opto} does not increase linearly with induction time; instead, signal increase tapers off after approximately 8 h of induction. After the end of the induction period, it takes approximately 6 h for luminescence to reach an uninduced steady state. The rate of decrease is reflective of the half-life of the luciferase protein rather than the off-time of the PULSE system expression: structural changes in EL222 and PhyB are expected to result in turning off of expression on a ms to s scale (*239, 240*). Normalised luminescence values did not decrease to pre-induction levels for lines #27 and 37 but remained slightly elevated at approximately 10 % their maximal level for at least 24 h after end of the induction.

Imaging of luminescence using a photon counting camera allows for a visualisation of signal distribution (Figure 4.2A). The luminescence signal was heterogeneously distributed in all PULSE lines with higher levels in the apical regions of the thalli. In some cases, no signal at all was detected at the centre of the thallus. The thalli of *M. polymorpha* develop from apical stem cells located in an invaginated 'notch' at the thallus extremity (*241*). As such the apical region is expected to grow faster than the basal region of the thallus. The processes involved in growth such as cell division or protein synthesis require ATP (*242*); the higher amounts of luminescence signal observed in apical tissue may reflect a combination of enhanced ability to accumulate large amounts of protein and higher availability of ATP to facilitate the conversion of D-luciferin to oxyluciferin.

4.3.2.2. Activity of PULSE in response to different light conditions

Three lines (#8, 27, 37) were taken forward for further characterisation. To determine the performance of PULSE in *M. polymorpha*, induction of expression in these lines in response to different intensities of red light (Figure 4.2B), to different ratios of red and blue light (Figure 4.2C), and to different combinations of red and far-red light (Figure 4.2D) were tested. Plants were grown from 36-day-old gemmae (red to blue light ratio), or from thallus cuttings grown for 1 week (red light intensity and different red/far-red light combinations experiments). In all cases the plants were placed in the tested light condition for 18 h before sampling.

Renilla luciferase is widely used as an internal control for luciferase-based assays. This allows for the normalisation of firefly luciferase measurements by calculating the ratio of firefly luciferase to the constitutively expressed *Renilla* luciferase activity, and

thereby reducing variability due to sample handling or expression levels (*243*). In PULSE, *Renilla* luciferase expression is under the control of the p*At*Ubi10 promoter. Unfortunately, no *Renilla* luciferase signal was detected in any experiments carried out in this work. This is probably due to lack of activity of the p*At*Ubi10 promoter in *M. polymorpha*. p*At*Ubi10 has been shown to lead to high expression levels in *N. benthamiana*, *Medicago truncatula*, *Lotus japonicus*, and *Hordeum vulgare* and is hence active in both monocotyledonous and dicotyledonous species. In addition, p*At*Ubi10 has been used in *P. patens* (*39*). The reason for its apparent inactivity in *M. polymorpha* is therefore unclear. The inability to normalise luminescence data has led to a high amount of variability in the data (Figures 4.2A-D). This variability probably represents different levels of luciferase expression between gemmae or thallus cuttings from the same line. A high level of variability between isogenic plants from the same line, both phenotypic and otherwise, has been observed throughout work undertaken for this thesis (as an example see Figures 5.1E and S5.1).

Normalisation is also carried out to allow for comparison of values across experiments. To enable this without a functioning normalisation module, the quantification of luciferase expression in each experiment was accompanied by a standard curve prepared from firefly luciferase. As such all data is expressed in terms of µg luciferase normalised by final plant weight.

Despite the high levels of variability, some conclusions can be made regarding the activity of PULSE in *M. polymorpha*. Firstly, there is no clear relationship between red light intensity and level of induction (Figure 4.2B). The level of expression does not increase linearly with the intensity of red light applied. Given this result, a red-light

intensity of 10 µmol photons m^{-2} s^{-1} was used for inducing expression in future experiments.

Secondly, testing different ratios of blue to red light showed that different lines have different rates of induction and of leakiness (Figure 4.2C). An induction rate of expression can be determined by calculating the difference in normalised luminescence between plants placed in monochromatic blue (expected to fully prevent expression) or red (expected to fully induce expression) light. Calculated this way, the induction rate was 20.6, 2.9, and 7.0 for lines #8, 27, and 37, respectively. The lower rate of induction in lines #27 and 37 is likely due to basal levels of expression through P_{Opto}, or 'leakiness'. In these lines, 10 µmol photons m^{-2} s^{-1} of blue light is unable to fully turn off expression. In addition, the lines responded with different sensitivity to the ratio of blue to red light; whereas lines #8 and 37 showed negligible leaky expression in the presence of 10 µmol photons m^{-2} s^{-1} of both blue and red light, in line #27 this light condition resulted in expression levels approaching those achieved with monochromatic red light.

Finally, basal expression levels in white light were characterized by quantifying the levels of luciferase expression in the long-day (LD) white light conditions normally used for plant growth, either with (W+FR) or without (W) supplemental far-red light (Figure 4.2D, see spectra in Figure S2.1). In these growth conditions, ratios of blue:red:far-red light are 4:2:1 and 8:4:1 in W+FR and W conditions, respectively. Both white light growth conditions were shown to lead to similar levels of expression as other conditions expected to prevent expression, such as dark and monochromatic far-red light conditions. Given these results, and the observation that PULSE transformants

grew better in white light supplemented with far-red light, W+FR was chosen for growth of these plants and will hereafter be referred to as the white light condition.

The data shown in Figures 4.2B and D was generated from plants grown from thallus cuttings, whereas the data shown in Figure 4.2C was generated from plants grown from gemmae. Propagation of *M. polymorpha* by thallus cuttings can lead to silencing (Susanna Streubel, formerly at the Gregor Mendel Institute, Vienna, Austria [personal communication]); this is visible in line #37 which responds less (Figure 4.2B) or not at all (Figure 4.2D) to monochromatic red light after thallus cutting subculture.

4.3.2.3. Exploration of induction strategies

To ensure adequate PHB production levels, it is likely that the expression of the PHB biosynthetic enzymes would need to be induced for a longer period than tested so far. To determine the long-term induction dynamics of PULSE in *M. polymorpha*, three different induction strategies were tested (Figures 4.2E-G). For these experiments, thallus cuttings were taken and allowed to grow for 1 week before being subject to the conditions shown.

Firstly, the response of PULSE to constant induction over 6 days was determined (Figure 4.2E). In lines #27 and 37, expression levels increased or stayed constant for 3 days after exposure to monochromatic red light; in line #8 expression increases only to day 1. For all lines tested, expression levels do not stay constant over the 6-day period, and instead start decreasing after day 1 or day 3 after induction for lines #8 or lines #27 and 37, respectively. For lines #8 and 27, expression is negligible after 6 days of exposure to red light.

Given an observation that PULSE transformants grew less well in white light without supplemental far-red light (W light conditions), it was hypothesized that this condition might be leading to basal expression levels through P_{Opto}, constituting a metabolic load and thereby negatively affecting plant growth. The data in Figure 4.2D showed that expression levels in W light are low or negligible. However, this represents a snapshot of expression 18 h after exposure to these conditions. A fuller picture of long-term expression dynamics shows that luciferase expression in white light without far-red increases to 3 days after exposure and then remains constant for at least another 3 days (Figure 4.2F). Expression levels are reduced compared to induction with red light but remain much more constant over the 6-day experimental period.

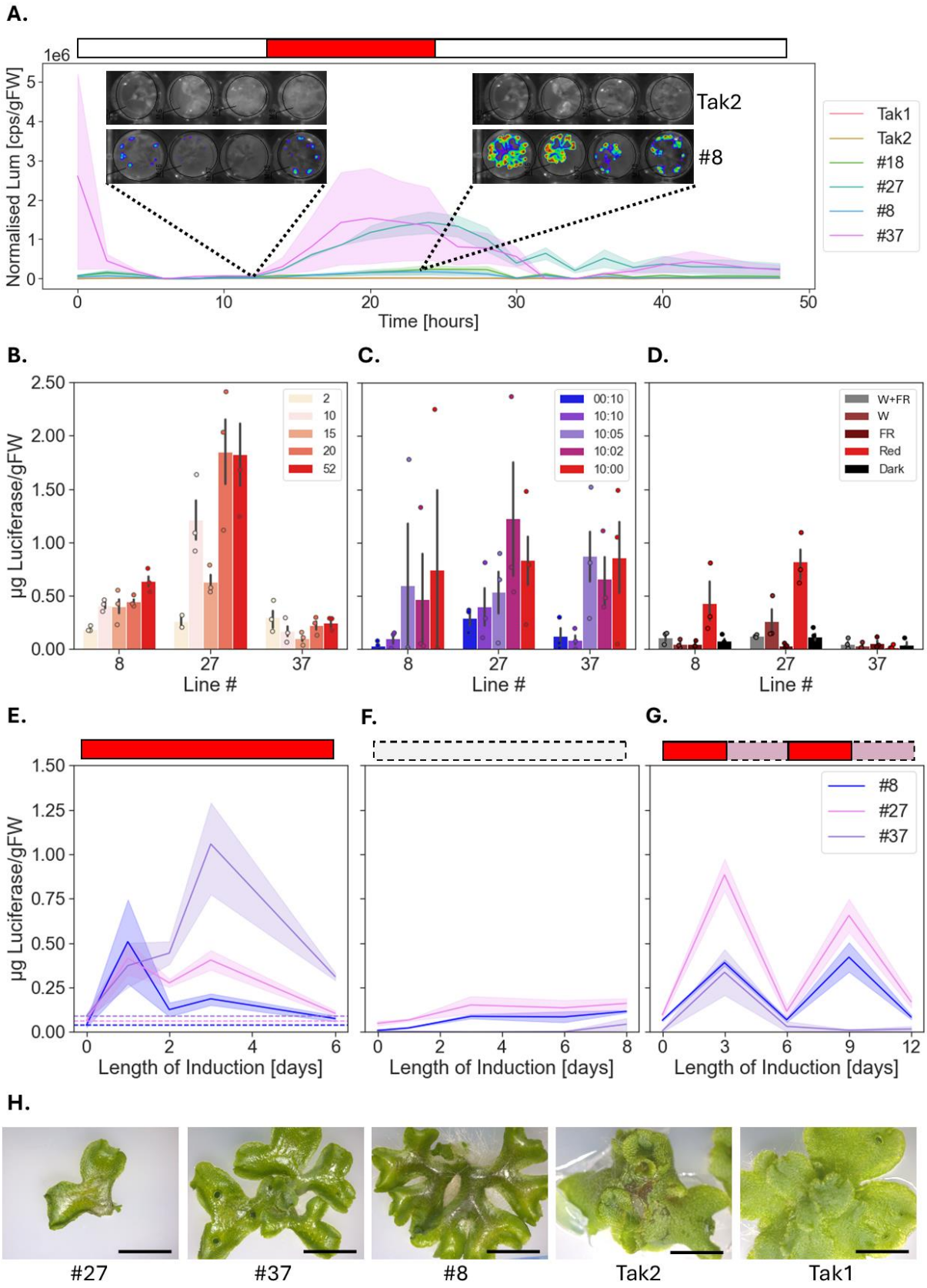
Finally, an intermittent induction scheme was tested (Figure 4.2G). Plants were placed in monochromatic red light for 3 days followed by W+FR conditions for 3 days, and this was repeated for a total of two cycles in each condition. Results show that this approach is effective at turning expression on and off, and that, apart from line #37, induction levels upon the second pulse of red light are comparable to the first. The lack of response of line #37 to the second pulse of red light, as well as the lack of induction shown in Figure 4.2F, are further evidence of silencing in this line.

4.3.2.4. Phenotypic effects of the PULSE system

Transformed *M. polymorpha* plants carrying PULSE display a notable developmental phenotype (Figure 4.2H). This was visible in spores several days after germination and could predictably be used to select plants carrying a functional PULSE system. The PULSE phenotype is not linked to induction as it occurs in plants both before and after exposure to monochromatic red light. The phenotype observed in PULSE

transformants includes slower growth, thinner and more branching thalli, increased cuticular deposition, reduced or no gemmae cup development, and low rates of gemmae germination. Some of these phenotypes could be partly rescued, as explained, by growing plants in the presence of far-red light.

Figure 4.2 (next page): Characterisation of the PULSE optogenetic system in *M. polymorpha*. (A) Dynamics and distribution of expression through PULSE in *M. polymorpha*. Counts per second (cps) normalised by final mass of each plant are shown in response to monochromatic red-light treatment; inlay represents raw data obtained from the photon counting camera, showing signal heterogeneity where red represents highest counts and blue lowest counts; n=4. Shaded region represents SEM. (B) Response of the PULSE system to different intensities of red light; intensities shown in units of µmol photons m^{-2} s^{-1}; n = 3. Error bars represent SEM. (C) Response of the PULSE system to different ratios of blue to red light; ratios of blue to red light intensities shown in units of µmol photons m^{-2} s^{-1}; n = 3. Error bars represent SEM. (D) Response of the PULSE system to different white light conditions; W+FR, white light with supplemental far-red; W, white light with no supplemental far-red; FR, far-red; n = 3. Error bars represent SEM. (E) Long term induction behaviour of PULSE over 6 days. Horizontal lines represent expression levels at t = 0; n= 3. Shaded region represents SEM. (F) Long term induction of PULSE in white light without the addition of far-red light; n= 3. Shaded region represents SEM. (G) Induction of PULSE in response to multiple 3-day pulses of red light; n = 3. Shaded region represents SEM. For E-G, a bar above the figure represents the light condition used where red is monochromatic red light, light grey is W, light purple is W + FR. (H) Phenotype of three PULSE lines and WT *M. polymorpha* Tak1 and Tak2 after 25 d of growth from gemmae in W+FR conditions. Scale bar is 5 mm.


A.
1e6
Normalised Lum [cps/gFW]
Time [hours]
Tak2
#8
Tak1
Tak2
#18
#27
#8
#37
B.
µg Luciferase/gFW
Line #
2
10
15
20
52
C.
00:10
10:10
10:05
10:02
10:00
D.
W+FR
W
FR
Red
Dark
E.
F.
G.
Length of Induction [days]
#8
#27
#37
H.
#27
#37
#8
Tak2
Tak1

4.4. Discussion

4.4.1. Challenges associated with working on an emerging testbed species

Although the body of work regarding *M. polymorpha* is expanding, tools and techniques for the genetic engineering of this species are still relatively undeveloped. During the establishment phase of a novel testbed species, it is inevitable that researchers will start working on the species for the first time, perhaps with no immediate support from researchers with direct experience of the species in question. In this context it would be very useful to have access to a database of information. In the case of *Marchantia*, recent developments have included an open registry of genetic parts (*244*), the OpenPlant *Marchantia* genetic toolkit and associated resources (*40*), and a catalogue of reporter proteins (*245*). Although very useful, none of these resources tackle basic issues somebody new to the field of *Marchantia* engineering might encounter. Issues identified in this work such as choice of transformation method, differences in transformation efficiency between Tak1 and Tak2 WT plants, insensitivity to kanamycin, inactivity of p*At*Ubi10, or silencing are known in the *Marchantia* community but are not published. A dedicated source of basic information regarding working with *M. polymorpha* would help establish this species as a novel testbed species for metabolic engineering.

4.4.2. An evaluation of PULSE in *M. polymorpha*

The activity of the PULSE system has previously been tested in *Arabidopsis*

protoplasts, in transiently transformed leaves of *N. benthamiana*, and in stably transformed *Arabidopsis* plants (*73*). In *Arabidopsis* protoplasts, induction rates up to 396.5-fold above those in darkness were observed. In transiently infiltrated *N. benthamiana* leaves, induction reached 28.7-fold 9 h after treatment with red light. Finally, two independent *Arabidopsis* lines had activation levels ranging from 10- to 21-fold. In *Marchantia*, induction rates reached 20.6-fold when measured as the difference of expression between monochromatic blue and red-light conditions. As such the induction rate of PULSE in *Marchantia* is broadly similar to published results in other species. Expression dynamics are also comparable between stable *Marchantia* and *Arabidopsis* lines; in both systems expression levels increase quickly upon induction with red light then level off (Figure 4.2A).

As well as induction, the two other important attributes of an optogenetic system are reversibility and leakiness. Reversibility of PULSE in *Marchantia* was demonstrated by the reduction in luciferase signal after induction when plants were placed in white light (Figures 4.2A and G). It takes approximately 6 h for luminescence levels to decrease to a steady-state uninduced level; this is faster than the approximately 20 h it takes for the same to happen in stable *Arabidopsis* lines (*73*).

In addition, in *Marchantia*, luminescence signal does not immediately decrease back down to a pre-induction baseline after exposure to red light (Figure 4.2A). This is the case also in stable *Arabidopsis* lines (*73*). However, luminescence eventually decreases to pre-induction basal levels as shown in plants after 3 days of white light (Figure 4.2G). In addition, some lines showed luciferase expression under blue or white light (Figures 4.2C and F). As such there is some expression through the promoter in

conditions other than monochromatic red light. Despite this, I can conclude that the PULSE system is fully functional in *Marchantia*.

4.4.3. Limitations of using PULSE in *Marchantia*

Although the PULSE system was successfully implemented in *M. polymorpha*, transgenic lines had a prominent developmental phenotype (Figure 4.2H). As explained in section 4.3.1, the PULSE system relies on the constitutive expression of three modules: R_{on}, B_{off}, and a normalisation element. It is possible that the phenotypic effects are due to the increased burden on the plant associated with constitutive protein expression. It is also possible, however, that they result from a specific interaction between a constitutively expressed module and native cellular processes in *Marchantia*. Given that phenotypic effects are present in plants grown only in white light, and are visible from the sporeling germination stage, they are likely due to one or more of the constitutively expressed PULSE construct proteins rather than FLuc which is only significantly expressed upon induction. No signal from the normalisation module was detected, an effect hypothesized to be due to inactivity of the p*At*Ubi10 promoter in *Marchantia*, and so phenotypic effects are unlikely to be due to expression of *Renilla* luciferase. In addition, although not tested in plants on its own, the expression of an EL222 and VP16-based optogenetic system in zebrafish embryos resulted in minimal morphological effects or toxicity (*246*).

It is likely that the developmental phenotype arises due to the constitutive expression of the R_{on} module; that it is derived from the plant proteins PhyB and PIF6 from *A. thaliana* makes it feasible that these proteins are interfering with developmental

processes in *Marchantia*. The *Marchantia* genome is very streamlined with no evidence of recent whole-genome duplications in its evolutionary history; this means that there is less genetic and regulatory complexity (*143*). For instance, in *Arabidopsis*, there are five different phytochromes (Phy) and eight PIFs (*247*). In *Marchantia*, however, there is only one gene each for a phytochrome (*Mp-Phy*) and its PIF (*Mp-PIF*) (*248*). The *Mp*Phy-*Mp*PIF pair encoded by these genes regulates a wide range of processes. For instance, *Mp*Phy and *Mp*PIF have been shown to be implicated in the germination of gemmae (*248*) and regeneration in thalli and sporelings (*249*). In addition, *Mp*PIF-mediated signalling has been shown to regulate meristem dormancy. The meristems that develop at thallus apices are responsible for the branching architecture in *Marchantia*, and their relative number, position, and activity determine the final shape of the thallus (*250*). The roles of *Mp*Phy and *Mp*PIF and the lack of redundancy in Phy-PIF signalling encoded in the *Marchantia* genome means that the expression of another Phy-PIF pair from *Arabidopsis* might be able to significantly interact with *Mp*Phy, *Mp*PIF, and/or their downstream targets, and that this might lead to the phenotypes observed. This is further supported by the fact that the deduced amino acid sequence of *Mp*Phy includes canonical land plant phytochrome domains and that a phylogenetic analysis of *Mp*PIF showed it to be closely related to the PIF family in *Arabidopsis* (*248*).

The developmental phenotype of PULSE *M. polymorpha* results in some experimental complexities. The propagation of *Marchantia* can be carried out either via gemmae or via thallus cuttings. One fully mature gemmae cup will contain approximately 100 gemmae, meaning the number of plants can rapidly and easily be scaled up (*235*). Given the inability and/or slowness with which the PULSE-carrying lines produced

gemmae cups, propagation for line maintenance and experiments was carried out by taking thallus cuttings. However, this resulted in silencing (see Figures 4.2D and F-G). This presents significant challenges to the long-term maintenance of PULSE lines and is a major limitation of using PULSE in this species. It should be noted that phenotypic defects were also noticed in stably transformed *Arabidopsis* lines, although the details of this have not been published (Uriel Urquiza-García, Heinrich Heine University, Düsseldorf, Germany [personal communication]).

4.4.4. Outlook for using inducible promoters in the metabolic engineering of *M. polymorpha*

Three examples of the use of inducible promoters in *Marchantia* have been published. The first is a heat shock responsive promoter based on the endogenous heat shock protein (HSP) gene *MpHSP17.8A1* which was shown to lead to inducible and transient expression upon heat shock at 37 °C. The *MpHSP17.8A1*-based promoter has been induced in whole plants as well as locally in response to laser-assisted heat treatment and used to drive the expression of several reporter proteins (*42, 251*). Additionally, small molecule-responsive promoters have been adapted for use in *Marchantia*. For instance, the oestrogen-inducible XVE system was used to drive expression of artificial microRNAs which were lethal when expressed constitutively (*230*). The XVE system is composed of a chimeric transcription factor built from the lexA DNA-binding domain from *E. coli*, the oestrogen-binding domain of the oestrogen receptor from *Homo sapiens*, and VP16 (*229*). In a different approach, the soybean auxin-inducible promoter *Pro*GH3 was used to detect auxin-mediated transcriptional activation *in planta* by monitoring GUS activity (*231*).

Both the heat-shock and small-molecule-mediated induction approaches have major drawbacks for applications in metabolic engineering. Heat shock is applied by transferring plates containing *M. polymorpha* thalli grown at 22 °C to 37 °C. However, plants displayed serious whole-plant tissue damage after heat treatments as short as 2 h (*251*). In contexts where induction is used to accumulate a product of interest, for instance, such short induction times may not be sufficient. For induction of the oestrogen-inducible promoter, plants were transferred to plates containing 5 μM β-oestradiol and response was seen 12 h after transfer. As such the induction of gene expression using the oestrogen-inducible promoter is not immediate and requires significant manual input. By contrast, the PULSE system presented in this work is not expected to have detrimental effects on plant health upon induction, and induction is easy and fast. Despite the developmental phenotype associated with PULSE, this system presents itself as a promising alternative to published inducible techniques in *M. polymorpha*.

Recently, another plant optogenetic system purported to be compatible with growth in white light has been reported (*232*). This system is based on the cyanobacterial CCaS-CCaR system, a green/red photoswitching transcription control system. By engineering CCaS to accept the phytochromobilin (PΦB) chromophore endogenously produced in plants, target gene expression was shown to be specifically repressed with blue light or blue-enriched white light, but active under other light regimes such as green-enriched white light. This system was tested in transiently transformed *N. benthamiana* leaves for the control over fluorescent protein, pigment, or immune response protein production. However, the system is considerably leaky and active in

dark conditions. Despite this, and given the ability to modulate activity by changing the spectral properties of white light, this system would warrant being tested in stable plant lines.

5. The production of PHB in *Marchantia polymorpha*

5.1. Introduction

PHB is a bioplastic precursor molecule produced by certain bacteria as a carbon-storage compound. The three enzymes responsible for PHB production in *R. eutropha* (see Figure 1.1A) have been engineered into plants with limited success, and no commercially viable levels of PHB accumulation in plants have so far been reported (see Table 1.1). One of the main obstacles to high levels of PHB production in plants is that it is almost always associated with negative effects on plant health. These effects depend on the subcellular location in which its production has been engineered but are thought to be intrinsically linked to acetyl-CoA availability. In the plastid, for example, the production of PHB competes with other acetyl-CoA-requiring pathways such as fatty acid and isoprenoid biosynthesis. This is accompanied by shading effects of PHB granules on thylakoids, leading to reduced photosynthesis and ATP starvation, as well as severely abnormal plastid morphology (*48, 52, 56, 63*).

In the cytosol, the production of PHB also competes with essential endogenous reactions for acetyl-CoA. Yields of PHB in the cytosol are generally lower than those achieved in the plastid, probably reflecting the reduced acetyl-CoA availability in this subcellular location. Indeed the highest published cytosolic yield to date in a terrestrial plant is 0.8 % per unit leaf dry weight (converted from value reported as %

per unit leaf fresh weight using conversion factor from (*202*))(*65*), but the highest plastidial yield is 40 % per unit leaf dry weight) (*52*). Despite this, in Chapter 3 of this work I used a stochiometric model of *Marchantia* central metabolism to determine that the cytosol was a more promising location for PHB engineering because (a) optimal flux paths to cytosolic PHB production did not include toxic or low-concentration metabolites, (b) the storage of citrate in the vacuole at night, which was subsequently used to produce PHB, mimics a known diurnal carboxylic acid storage pattern, (c) it avoids physical effects of PHB granules in the plastid such as thylakoid shading and plastid deformation, and (d) ACL activity emerged as a potentially straightforward engineering target. The model was used to identify specific engineering targets which might increase the capacity of the existing metabolic network to support maximal PHB production. This approach contrasts with the vast majority of plant PHB-engineering approaches carried out to date which have relied on the sole introduction of the three enzymes already mentioned. In this chapter these computationally identified targets will be implemented in *M. polymorpha* and their effect on PHB yield characterised.

As well as a consideration of the capacity of the entire network to support PHB production, inducibility can also be used to decrease the effect of an engineering intervention on endogenous metabolism. By controlling when PHB production occurs, it can be at least partly temporally separated from essential processes such as plant growth. Inducible promoters have been used to drive PHB production in plastids and shown to help reduce negative phenotypic effects such as growth deficiency and male sterility (*70-72*). To my knowledge, inducible promoters have not been used for

cytosolic PHB production. In Chapter 4 I showed that the PULSE optogenetic system is fully functional in *M. polymorpha*. In this chapter, PULSE will be tested for its ability to control cytosolic PHB production.

To evaluate the effect of the computationally derived targets and optogenetic induction on yield of PHB and plant health I also generated *M. polymorpha* lines constitutively expressing the three enzymes required for PHB synthesis. These were used as a baseline against which the effects of further engineering could be evaluated.

PHB production in plants presents challenges in terms of substrate availability, low product titre, and complex interactions with endogenous metabolism, issues which are generally applicable to plant metabolic engineering. The approaches presented in this chapter have value in presenting solutions for the engineering of many types of molecules, as well as furthering the tools available for using *Marchantia* as a testbed species in metabolic engineering.

5.2. The constitutive production of PHB in *M. polymorpha*

5.2.1. The design of a polycistronic construct for the constitutive expression of β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase

In order to drive the constitutive expression of β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase I designed two versions of a polycistronic construct. Briefly, the coding sequences of the three proteins with stop codons removed were separated by self-cleaving viral peptide sequences. To aid the purification of the

enzymes from the crude cell lysate, 10×His tags were added to the N-terminus of all three coding sequences. For protein detection a c-Myc, FLAG, and HA tag was added to β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase, respectively, on the C-terminus (β-ketothiolase and PHB synthase) or on the N-terminus (acetoacetyl-CoA reductase).

Two different self-cleaving viral peptide strategies were tested. 2A peptides are 18-22 amino acid viral oligopeptides which mediate the cleavage of polypeptides during translation in eukaryotic systems, by the ribosome skipping of the formation of a glycyl-prolyl peptide bond at its C-terminus (*177*). Different viral 2A peptides have been named after the virus in which they originated, such as F2A (from the foot-and-mouth disease virus), P2A (from porcine teschovirus-1), and T2A (from *Thosea asigna* virus). 2A peptides have been shown to lead to relatively high levels of protein expression, and their small size minimizes the chance of interfering with protein function (*252*).

For the first construct, P2A and T2A sequences were placed at the junction between the β-ketothiolase and acetoacetyl-CoA reductase fusion proteins, and the acetoacetyl-CoA reductase and PHB synthase fusion proteins, respectively. These two 2A sequences were chosen due to their established high cleavage efficiency (*177, 253*), and data showing multiple identical 2A peptides should not be used in the same construct (*252*). A glycine-serine-glycine (GSG) linker was added on the N-terminus of both 2A peptides, shown to improve cleavage efficiency (*178*).

The second approach involved the IntF2A self-cleaving fusion protein domain, composed of a *Ssp* DnaE mini-intein variant engineered for hyper-N-terminal autocleavage which is covalently linked to the F2A peptide (*179*). This domain was

placed between protein coding regions in the same way as P2A and T2A to create a self-excising fusion protein domain.

Both versions of the construct were cloned into the expression vector p*Mp*GWB103 specifically designed for overexpression in *M. polymorpha (42)*. This vector places the coding sequence downstream of the strong endogenous p*Mp*EF1α promoter and carries hygromycin resistance.

5.2.2. Characterisation of constitutive PHB production in *M. polymorpha*

M. polymorpha sporelings were transformed with the constitutive PHB constructs. Preliminary analysis found detectable PHB levels only in plants transformed with the InteinF2A version of the construct and hence these plants were taken forward for further analysis. 22 lines were grown for 29 d from gemmae and screened for PHB yield and growth rate (Figures 5.1A and B). PHB is expressed in terms of % PHB per g fresh weight (%PHB gFW^{-1}) as this is the predominant unit used in the literature. The 22 lines showed a large range of PHB yields and of phenotypes (see Figure S5.1). The most productive lines produced PHB levels similar to previous yields in *Arabidopsis,* which also represents the highest yield of cytosolic PHB published to date (*65*). There was a strong negative correlation ($P = 0.0014$) between average final size of each line in mm^2 after 29 d of growth and PHB yield. Thallus area is a good proxy for size in *Marchantia* given its predominantly 2-dimensional growth (*254*). The negative correlation supports both the computational finding that the accumulation of biomass and of PHB are in direct competition (Figure S3.1) and previous work reporting negative growth

phenotypes of plants producing cytosolic PHB (Table 1.1).

Three of the highest yielding lines were taken forward for further investigation (lines #3, 18, and 25) and were grown for 39 d from gemmae. These lines had diverse phenotypes (Figure 5.1C) and were smaller than WT plants (Figure 5.1B and S5.2). The size of lines #3, 18, and 25 decreased by 78, 93, and 74 % compared to Tak1 plants and 46, 84, and 35 % compared to Tak2 plants, respectively. All three lines had visible gaps in the dorsal epidermis not observed in WT plants, which were hypothesized to be enlarged air pores. This was most obvious in line #18 (Figure 5.1C). PHB yields comparable to published values were confirmed in these lines (Figure 5.1E) as was the expression of all three enzymes responsible for the biosynthesis of PHB (Figure 5.1F). To aid in their detection, His-tagged fusion proteins as described in section 5.2.1 were Ni-affinity purified from the crude plant lysate before SDS-PAGE and Western blotting. I was unable to detect the FLAG-fused acetoacetyl-CoA reductase using an α-FLAG antibody, but this enzyme was successfully detected using a α-PhbB antibody kindly gifted by Yves Poirier (Université de Lausanne, Switzerland). This was probably due to cleavage between the FLAG-tag and the protein coding sequence, which was confirmed given the detection of acetoacetyl-CoA reductase in the crude lysate but not in the elution from the Ni-affinity column (the FLAG tag was downstream of the His tag in the construct). Using Nile Blue A staining I was also able to visualize the PHB granules in gemmae of the transformed lines (Figure 5.1D). These granules were approximately 5 µm in diameter and were never observed in untransformed *M. polymorpha* plants. PHB granules in *R. eutropha* and other PHB-accumulating bacteria are usually 0.2-0.5 µm in diameter (*255*), whereas granules produced in plants

are larger with a maximal diameter of 10 µm (*51*) but are usually smaller than 1 µm (*56, 65*). The size of these granules is thus in line with previous reports of PHB in plants.

There was considerable variability in the yield of PHB obtained from isogenic plants grown from gemmae (Figure 5.1E). In line #3, for instance, yields varied between 0.014 and 0.17 %PHB gFW^{-1}. Differences in performance on the individual plant level have been previously reported for plants producing PHB (*201*).

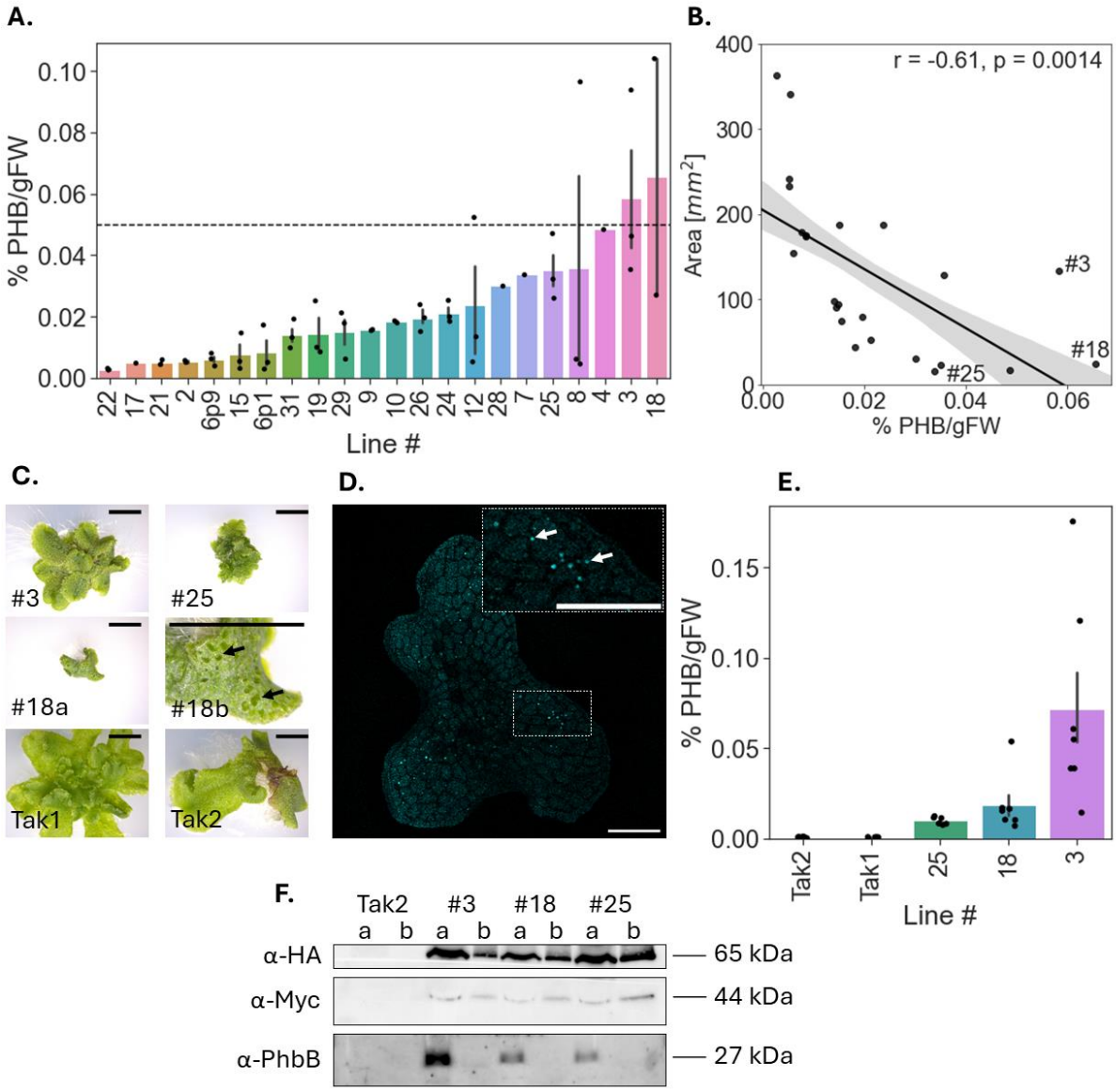


Figure 5.1: The characterisation of *M. polymorpha* lines constitutively producing PHB. (A) 22 lines were screened for PHB production in terms of %PHB gFW^{-1}. Dotted line represents highest published cytosolic PHB yield in a plant. n = 1-3. Error bars indicate SEM. (B) The relationship between PHB production and final plant size in mm^2 after 29-d growth in the 22 screened lines. Average values for each line are plotted. Shaded area represents SEM. Pearsons correlation coefficient (r) and *P*-value (p) shown in top-right corner. (C) Images of three selected high PHB-producing lines and WT *M. polymorpha* after 29 d of growth. Scale bar is 5 mm. Black arrows indicate enlarged pore-like structures. (D) Representative image of Nile Blue A-stained PHB granules in a gemma of line #25. Granules indicated in inset with white arrows. Scale bar is 100 µm. (E) Confirmation of PHB yield in three selected high-PHB producing lines grown for 39 d from gemmae. n = 6. Error bars represent SEM. (F) Western blot for the detection of β-ketothiolase (α-Myc), acetoacetyl-CoA reductase (α-PhbB), and PHB synthase (α-HA). Detection in crude lysate (a) and after elution from Ni-affinity column (b) shown. Loading is normalised based on amount of plant starting material.

5.3. Inducible control over PHB accumulation using an optogenetic system

The PULSE system as explained in Chapter 4 was modified to allow for optogenetic control over PHB production. Briefly, the coding sequence of firefly luciferase previously used as a reporter was replaced with the coding sequence of all three PHB-producing enzymes separated by InteinF2A as described in section 5.2.1 such that the expression of the polycistronic construct was under the control of P_{Opto}. The modified PULSE system was transformed into *M. polymorpha* sporelings. Inducible PHB lines were grown in long-day white light conditions supplemented with far-red light (W+FR), based on observations that the PULSE transformants grew better in the presence of far-red light (see section 4.3.2.2). As with the PULSE lines induction of expression was achieved by transferring plants to a chamber at 22 °C with 10 µmol photons m^{-2} s^{-1} monochromatic red light.

5.3.1. The characterisation of *M. polymorpha* lines inducibly producing PHB

23 inducible PHB lines were screened for their ability to produce PHB in response to red light. Briefly, lines were grown from gemmae for 16 d before three three-day induction periods separate by three-day periods in normal growth conditions. At the end of the 31-day period, samples were collected and PHB extracted. Results of this screen are shown in Figure 5.2A. As with the constitutive lines, inducible PHB lines showed variable phenotypes (Figure S5.1). From this analysis it was clear that only three lines accumulated significant levels of PHB at the end of the experimental period

(lines #43, 88, 112). The PHB levels were comparable to those in the constitutive lines (Figure 5.1). The relationship between final plant size in mm^2 and yield of PHB was investigated. There was a moderate negative relationship ($P = 0.053$) between final size and PHB accumulation in the inducible PHB lines tested (Figure 5.2B).

Inducible PHB lines had phenotypes reminiscent of both PULSE lines (Figure 4.2H) and constitutive PHB lines (Figure 5.1C). Some lines showed branching phenotypes as observed in PULSE lines (Figure 5.2C); these lines produced non-viable gemmae and so were not part of the screen or taken forward for further analysis. Lines #43, 88, and 112 showed both an increase in cuticular deposition as observed in PULSE transformants as well as the gaps in the dorsal epidermis as seen in plants constitutively producing PHB. These gaps were only seen in inducible lines after induction. In addition, they also produced reduced numbers of gemmae cups.

Lines #43, 88 and 112 were taken forward for further analysis. The three lines were grown from gemmae for 30 d with evenly timed bursts of induction using red light as shown in Figure 5.2F. Some plants were grown for the same amount of time but without induction, and samples taken at the end of the growth period (i.e. immediately after the fifth 3-day induction period for the samples with induction) were probed with antibodies for the detection of β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase to investigate the effect of induction on the protein level (see Figure 5.2E). β-ketothiolase was not detected in samples from line #88; this is most probably a technical problem given that both acetoacetyl-CoA reductase and PHB synthase were detected, and the proteins are expressed as part of a polycistronic construct (and that significant levels of PHB are detected in this line after induction). In all other cases, the

levels of the three enzymes increased significantly after induction, showing that the P_{Opto} promoter was being successfully activated by monochromatic red light. There were variable levels of expression in white light, showing the system to be somewhat leaky. As expected, no protein was detected in WT Tak1, and all three proteins were detected in constitutive PHB line #3 included as a positive control.

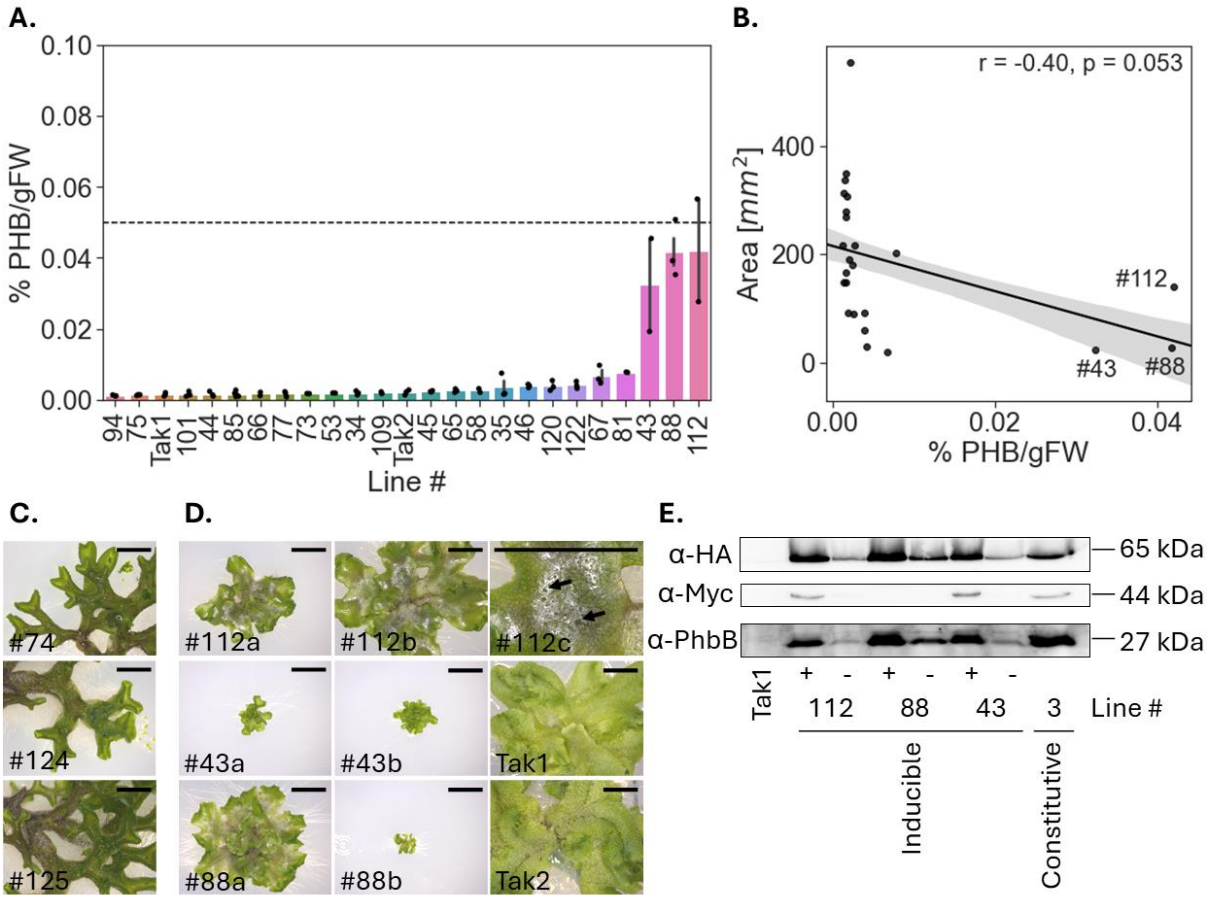
As well as protein extraction and detection, plants grown as explained were also used to confirm PHB production levels following induction (Figures 5.2G and H). Unfortunately, due to the effects on gemmae cup production as described, line #43 was excluded from this analysis. To allow for a direct comparison of PHB production levels between constitutive and inducible PHB lines, the three constitutive lines as previously described as well as inducible PHB lines #112 and 88 were grown from gemmae according to the induction scheme shown in Figure 5.2F. When expressed in terms of %PHB gFW^{-1} the inducible lines outperform the constitutive lines in these conditions with 2-3-fold higher yields (Figure 5.2G).

I hypothesized that inducible control over PHB production might temporally separate growth and PHB production and thereby reduce the negative growth phenotypes associated with PHB production and seen in the constitutive PHB lines. If this was the case, we would expect to see larger plants given the same levels of PHB accumulation. However, inducible PHB lines are roughly the same size (#112) or smaller (#88) than constitutive PHB lines grown in the same conditions (Figure S5.2A). For this reason, when reported as total amount PHB per plant, there is no longer a clear advantage to the inducible strategy (Figure 5.2H). As such, the PULSE optogenetic system can be used to drive PHB production to levels similar to a constitutive promoter, but under the

conditions tested did not help alleviate effects on plant growth associated with PHB accumulation. This could also be related to the effects on plant phenotype observed with the PULSE system on its own (Figure 4.2H).

It became clear during this analysis that the PULSE-induction growth conditions being used were influencing plant growth. WT and constitutive PHB lines grown with three three-day bursts of red light were significantly larger than plants grown only in white light (Figure S5.2A). In addition, constitutive lines were grown in white light not supplemented far-red light. As such it was either the presence of far-red light or the induction itself which was causing this increase in growth. In order to differentiate between these two scenarios, the size of WT Tak1 and Tak2 plants was investigated before and after three bursts of induction. Plants were grown for 17 d in white light with and without supplementary far-red; plants grown with far-red were then induced in three bursts as described until all plants were 31 d old. Plant area was measured on day 17 and on day 31. Results show that at day 17 there is no significant difference between Tak2 plants grown in white light with or without far red, whereas Tak1 plants grown with far-red were smaller (Figure S5.2B). At day 31, however, plants having been placed in induction conditions were significantly larger than those kept in white light growth conditions (Figure S5.2C). This shows that it is probably the three-day periods of monochromatic red light which are affecting plant growth. Indeed, it is well documented that growth of plants in monochromatic red light causes phenotypes such as increased plant height and leaf area (*256*). More alarmingly, plants grown in monochromatic red light also have reduced photosynthetic capacity known as the 'red light syndrome' (*257*).

Figure 5.2 (next page): The characterisation of *M. polymorpha* lines inducibly producing PHB. (A) 23 lines were screened for PHB production in response to monochromatic red light. Dotted line represents highest published cytosolic PHB yield in a plant. n = 2-3. Error bars represent SEM. (B) The relationship between PHB production and final plant size after 31-day growth in the 23 screened lines. Average values for each line are plotted. Shaded area represents SEM. Pearsons correlation coefficient (r) and *P*-value (p) shown in top-right corner. (C) Images of three inducible PHB lines with branching phenotypes. (D) Images of three selected high inducible PHB-producing lines and WT *M. polymorpha* after 31 d of growth and induction. Scale bar is 5 mm. Black arrows indicate enlarged pore-like structures. (E) Western blot for the detection of β-ketothiolase (α-Myc), acetoacetyl-CoA reductase (α-PhbB), and PHB synthase (α-HA) both post (+) and without (-) induction in crude plant lysate. Loading is normalised based on amount of plant starting material. (F) Induction scheme used for further characterisation of inducible PHB lines #43, 88 and 112. Sampling point indicated with a black arrow. (G) PHB levels expressed as %PHB gFW^{-1} in both constitutive (#18, 25, 3) and inducible (#112, 88) lines grown according to the scheme shown in F. n = 6. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U). (H) PHB levels expressed as total mg PHB per plant in both constitutive (#18, 25, 3) and inducible (#112, 88) lines grown according to the scheme shown in F. n = 6. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U).


A.
% PHB/gFW
Line #
94 75 Tak1 101 44 85 66 77 73 53 34 109 Tak2 45 65 58 35 46 120 122 67 81 43 88 112
B.
r = -0.40, p = 0.053
Area [mm²]
% PHB/gFW
#112
#43
#88
C.
#74
#124
#125
D.
#112a
#112b
#112c
#43a
#43b
Tak1
#88a
#88b
Tak2
E.
α-HA
α-Myc
α-PhbB
65 kDa
44 kDa
27 kDa
Tak1
112 88 43 3 Line #
Inducible
Constitutive
F.
Induction scheme
3 days, W+FR
G.
% PHB/gFW
**
18 25 3 112 88
Line #
H.
Total PHB per plant [mg]
n.s.
18 25 3 112 88
Line #

5.3.2. The dynamics of PHB accumulation in constitutive and inducible PHB-producing plants

In order to gain a fuller understanding of the interplay between growth and PHB accumulation, and how this varied between inducible and constitutive expression scenarios, I monitored the accumulation of PHB and growth of plants over time. Due to reduced gemmae cup development in inducible lines, thallus cuttings were used for this experiment. Cuttings were taken from all lines and grown according to the induction scheme shown in Figure 5.3C. Samples were taken prior to the first three-day period of induction, as well as after each induction period. Data was expressed both in terms of %PHB gFW^{-1} (Figure 5.3A) and total amount of PHB per plant (Figure 5.3B). For the inducible lines, some thallus cuttings were never exposed to induction conditions but grown the entire period in white light, as an additional control for the proper functioning of the PULSE optogenetic system.

There was very minimal production of PHB in thallus cuttings grown for the full growth period without induction, showing that the small amounts of protein expression detected without induction (Figure 5.2E) are insufficient to sustain significant PHB synthesis. Interestingly, however, significant levels of PHB were detected in thallus cuttings grown for 3 d without induction (0 d of induction). Thallus cuttings are taken from the apical part of the thallus. Spatial imaging of luciferase expression in the PULSE system showed that the highest levels of expression are obtained in the apical region of the thallus (Figure 4.2A). As such it seems reasonable to assume that the highest levels of leaky expression might also occur in the apical thallus tissue.

The dynamics of PHB accumulation over time varies considerably between the constitutive and the inducible PHB-producing lines. In terms of %PHB gFW^{-1}, in the constitutive lines the levels of PHB stay constant or decreases over the experimental period (Figure 5.3A). This contrasts with inducible lines where PHB levels increase over the experimental period, albeit with different accumulation patterns across the three lines tested. In general, the final levels of PHB expressed as %PHB gFW^{-1} are higher in the inducible lines than in the constitutive lines, echoing results already achieved when growing plants from gemmae (Figure 5.2G).

As explained, expressing PHB yields as total mg of PHB per plant can capture the interplay between PHB production and plant growth. As such the total mg of PHB produced per plant was monitored over time for both constitutive and inducible PHB-producing plants (Figure 5.3B). Interestingly, whereas when grown from gemmae no apparent benefit is seen for an inducible PHB-producing strategy (Figure 5.2H), there is a clear increase in total mg of PHB in inducible lines #88 and 112 compared to constitutive lines #3, 18, and 25 when grown from thallus cuttings (Figure 5.3B).

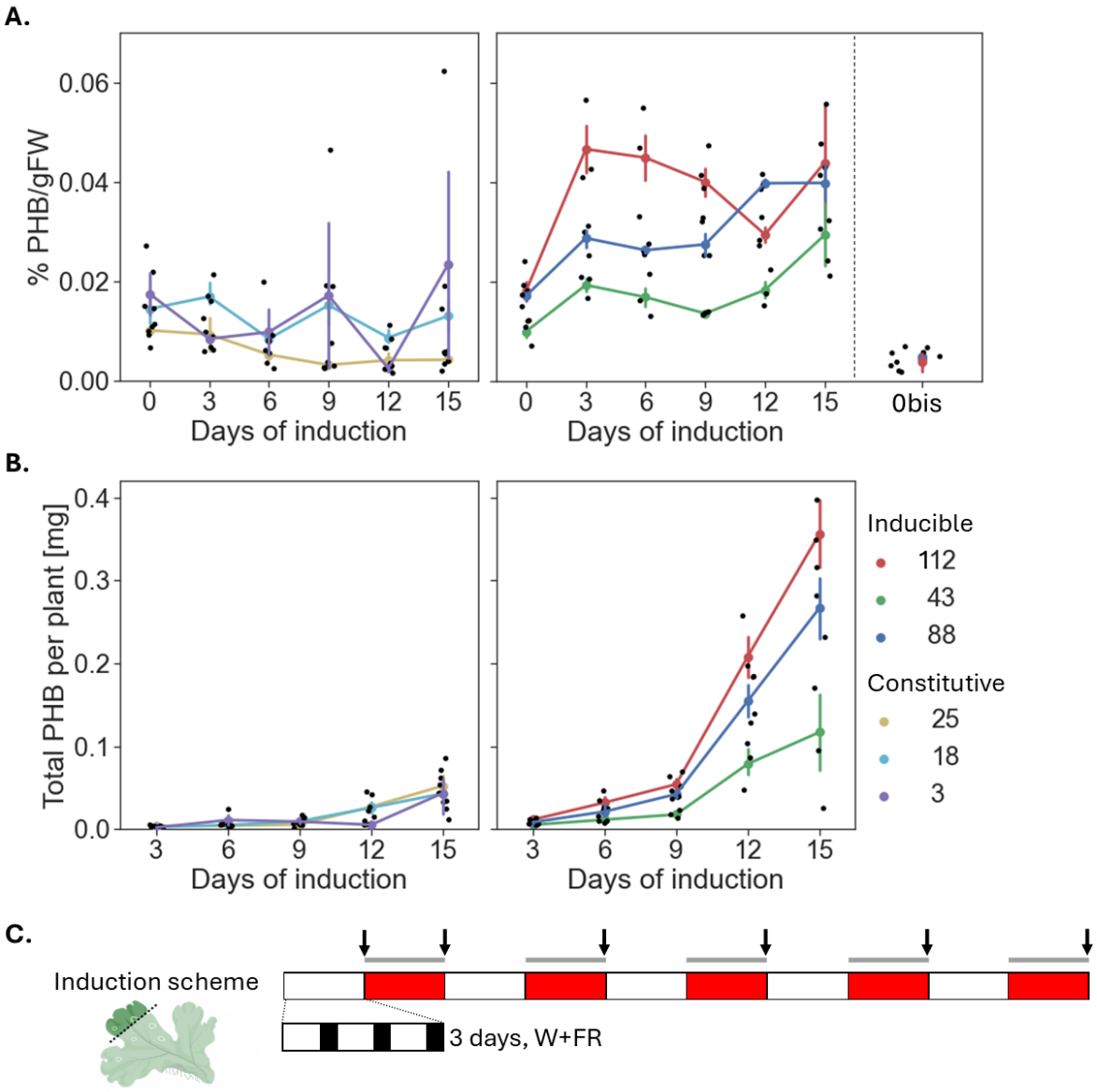


Figure 5.3: The dynamics of PHB accumulation in constitutive and inducible PHB-producing plants. (A) The change in PHB contents expressed in terms of %PHB gFW^{-1} of thallus cuttings from both constitutive and inducible PHB-producing plants over time. PHB levels in inducible lines grown for the same amount of time but not induced are also shown (0bis). (B) The change in PHB contents expressed as total mg per plant of thallus cuttings from both constitutive and inducible PHB-producing plants over time. Legend applies to both (A) and (B) and error bars represent SEM in all conditions. n = 3 in all conditions. (C) Induction scheme used in this experiment. Sampling times are indicated using black arrows.

5.4. Ancillary metabolic targets for increasing PHB accumulation in *M. polymorpha*

5.4.1. The design of a metabolic engineering strategy to increase cytosolic PHB accumulation in *M. polymorpha*

In Chapter 3 a computational model of the central metabolism of *M. polymorpha* was developed and was interrogated for its ability to support the synthesis of PHB alongside the accumulation of biomass. Analysis of the model identified possible engineering targets, on top of the introduction of the three PHB enzymes, for the production of cytosolic PHB (so-called 'ancillary' targets), and these are outlined in Table 3.4. The final selection of ancillary engineering targets was necessarily based on computational findings as well as experimental considerations and will be described here.

The analysis of the model revealed that the flux through ACL was the main differentiator between biomass-only and PHB-production scenarios. High flux through ACL ensures sufficient acetyl-CoA provision for both biomass accumulation and the first step in PHB synthesis. The human version of ACL (ACLY) is homotetrameric (*258*), in contrast to the heterooctameric plant version (*111*), which facilitates cloning and heterologous expression. ACLY is allosterically regulated through three key lysine residues (K540, K546, K554) shown to be ubiquitinated and thus targeting the enzyme for proteosomal degradation. The mutation of these three residues to glutamine (3KQ) or to arginine (3KR) resulted in a significant increase in protein stability and acetyl-CoA levels in human cancer cell lines (*259*). As the acetylation-mimetic 3KQ mutations

resulted in the highest acetyl-CoA accumulation levels this version of ACLY was selected for heterologous expression, and henceforth referred to as ACLY3KQ.

As an alternative to ACLY3KQ, I also designed a version of the construct carrying a chimeric ACL protein generated by linking the ACLA1 and ACLB2 subunits from *A. thaliana* ACL with a linker region from *Drosophila melanogaster* ACL, henceforth referred to as *At*ALB. The overexpression of this chimeric ACL has previously been shown to rescue growth phenotypes associated with constitutive PHB production in *A. thaliana (61)*.

Selection of other targets was hampered by a lack of molecular knowledge of the proteins involved. Because of time constraints, I only chose targets where target overexpression had previously been demonstrated to lead to increased levels of the desired molecule. Although *Cs*Cit1, a vacuolar citrate/proton symporter from *Citrus sinensis (260)*, was a promising target for increasing vacuolar citrate accumulation, the lack of any experimental knockout/overexpression evidence precluded its choice as a target. Knockouts of both mitochondrial malate importer/citrate exporter DIC2 (*212*) and cytosolic fumarase FUM2 (*206, 261*) have been generated, but no overexpressing lines characterized. Although the overexpression of *At*ALMT9 has not been carried out to my knowledge, the overexpression of the closely related *Sl*ALMT9 was shown to increase malate contents in tomato (*262*). Similarly, the overexpression of *At*tDT resulted in increased malate and reduced citrate contents in *Arabidopsis*, presumably due to export of citrate from the vacuole and subsequent catabolism of citrate in the cytosol (*181*). Given that an increase in malate and a decrease in citrate storage in the vacuole were both shown to be advantageous in terms of cytosolic PHB

production (see Chapter 3), *At*tDT was also selected as a target.

The coding sequences of ACLY3KQ or *At*ALB and *At*tDT with stop codons removed were separated by the coding sequence of the InteinF2A self-cleaving viral peptide. To aid in detection using western blotting, a c-Myc tag was added to the C-terminus of ACLY3KQ and *At*ALB, and a FLAG tag added to the N-terminus of *At*tDT. The constructs were cloned into the expression vector p*Mp*GWB303 specifically designed for overexpression in *M. polymorpha (42)*. This vector places the coding sequence downstream of the p*Mp*EF1α promoter and carries chlorsulfuron resistance.

5.4.2. The effect of the overexpression of ACL and *At*tDT on the yield of PHB in constitutive PHB-producing lines

Any of the conclusions made regarding the effect of the ancillary constructs have to be tempered by the fact that neither protein expression nor activity have yet been confirmed due to time constraints. As with all other transgenic lines the integration of the correct DNA into the plant genome was confirmed, but protein expression should be ascertained using a western blot and enzyme activity, at least of ACLY3KQ and/or *At*ALB, should be characterised as in (*61*).

Both versions of the ancillary target constructs were super-transformed into constitutive PHB lines #3, 18, and 25 using thallus transformation. Despite many attempts, no successful transformants were obtained for constitutive PHB line #3. For lines #18 and 25, transformation was more successful with the ACLY3KQ version of the construct than the *At*ALB version. 10 and 5 lines of line #18 super-transformed with ACLY3KQ- or AtALB-containing constructs, respectively, were screened for PHB

accumulation levels. For line #25, 11 and 3 lines were screened for both construct versions, respectively. Lines were grown from gemmae for 31 d under normal conditions, growth monitored, and PHB extracted. For comparison constitutive lines #18 and 25 were grown alongside the ancillary intervention lines.

The screen revealed a diverse range of responses to the ancillary intervention constructs (Figure 5.4A). Five ancillary lines showed higher levels of PHB accumulation than its parent line #18 (lines #18ACLY9, 4, 8, 6, 10); no lines with increased PHB levels were obtained when constitutive line #18 was super-transformed with the *At*ALB version of the construct. For line #25, two ancillary lines with higher PHB levels were obtained (lines #25ACLY10, 11). Again, no lines with increased PHB levels were obtained when constitutive line #25 was super-transformed with the *At*ALB version of the construct. To gain an understanding of the impact of the ancillary interventions on both growth and PHB accumulation, a correlation analysis was undertaken between the two variables (Figure 5.4B). There was a strong negative correlation ($P = 0.000032$) between final plant area and accumulated levels of PHB.

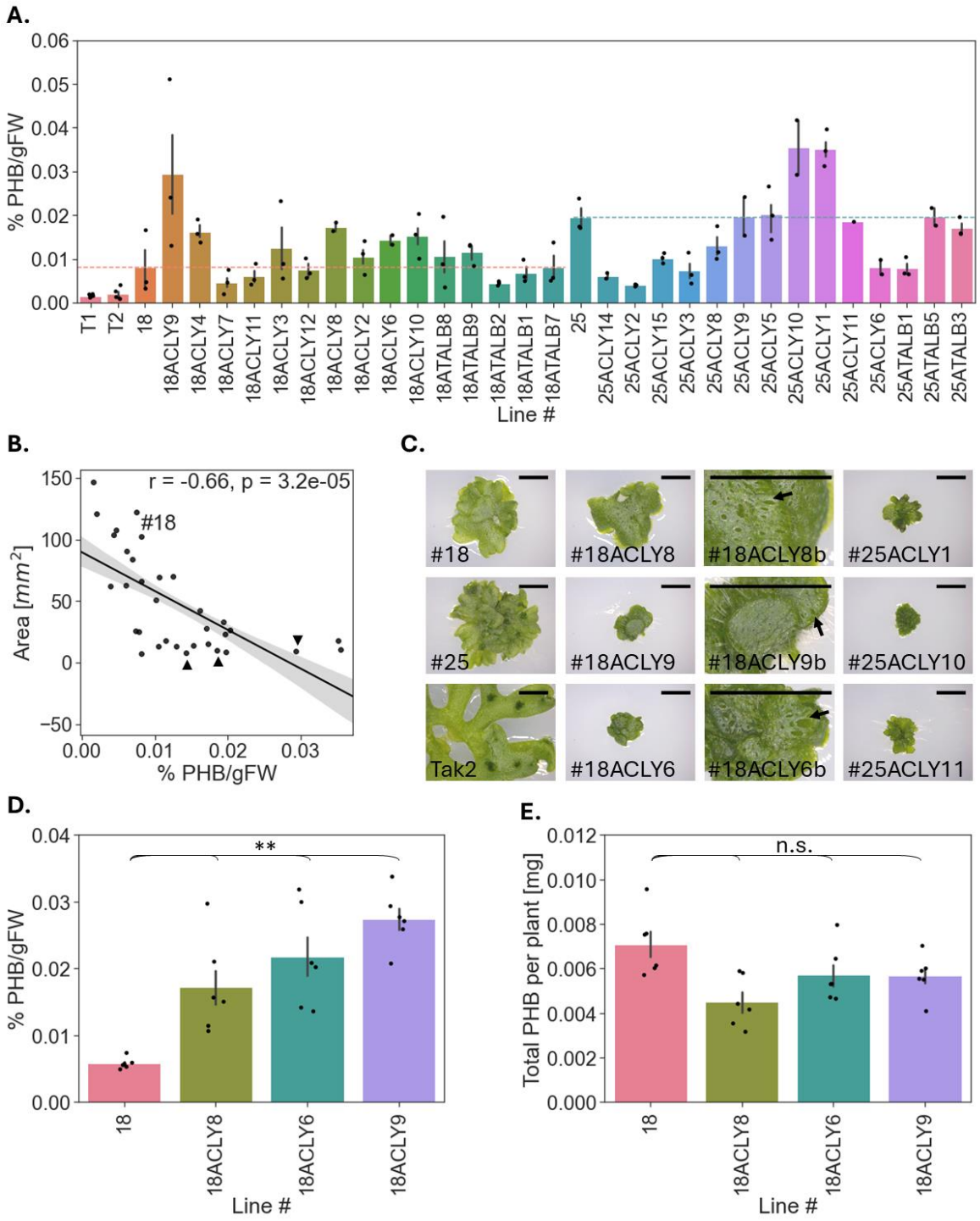
A range of phenotypes was observed in ancillary target lines derived from both constitutive PHB lines #18 and 25 (Figure S5.1). In general, the most severe phenotypes were associated with the highest levels of PHB production and included dramatically reduced size (Figures 5.4B and 5.4C). The large openings observed on the dorsal side of thalli in both constitutive PHB lines (Figure 5.1C) and inducible PHB lines (Figure 5.2D) were also observed in some ancillary lines and were most obvious in lines derived from constitutive line #18.

Based on the results of the screen, lines #18ACLY6, 18ACLY8, AND 18ACLY9 were

taken forward for further analysis. Plants were grown from gemmae for 31 d, growth monitored, and PHB extracted at the end of the growth period. When expressed in terms of %PHB gFW^{-1}, PHB yields in lines carrying ancillary interventions were up to 4.7-fold higher than in line #18 (line #18ACLY9, Figure 5.4D) (P = 0.0022). However, given the smaller size of these lines, when expressed in terms of total mg of PHB accumulated per plant the advantageous effect of this further engineering intervention was ablated (Figure 5.4E).

As such the introduction of the computationally derived targets was able to increase the production of PHB when expressed in terms of %PHB gFW^{-1}. However, this was associated with negative effects on biomass accumulation, due to either the increased levels of PHB or negative effects of the interventions themselves on *M. polymorpha*.

Figure 5.4 (next page): The effect of the overexpression of ACLY3KQ or *At*ALB and *Att*DT on the yield of PHB in constitutive PHB-producing lines. (A) A total of 29 lines transformed with the ancillary construct were screened for PHB production. Dotted line represents yield of PHB obtained in constitutive PHB lines #18 (pink) and 25 (blue). n = 2-3. Error bars represent SEM. (B) The relationship between PHB production and final plant size after 31-day growth in the 29 screened lines. Arrows represent values for (from left to right) #18ACLY8, 18ACLY6, and 18ACLY9. Average values for each line are plotted. Shaded area represents SEM. Pearsons correlation coefficient (r) and *P*-value (p) shown in top-right corner. (C) Images of constitutive PHB-producing lines #18 and 25 and selected high PHB-producing ancillary construct-carrying lines derived from them. Scale bar is 5 mm. Black arrows indicate enlarged pore-like structures. (D) PHB levels expressed as %PHB gFW^{-1} in both constitutive (#18) and ancillary (#18ACLY6, 18ACLY8, 18ACLY9) lines. n = 6. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U). (E) PHB levels expressed as total mg PHB per plant in both constitutive (#18) and ancillary (#18ACLY6, 18ACLY8, 18ACLY9) lines. n = 6. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U).


A.
% PHB/gFW
0.06
0.05
0.04
0.03
0.02
0.01
0.00
T1
T2
18
18ACLY9
18ACLY4
18ACLY7
18ACLY11
18ACLY3
18ACLY12
18ACLY8
18ACLY2
18ACLY6
18ACLY10
18ATALB8
18ATALB9
18ATALB2
18ATALB1
18ATALB7
25
25ACLY14
25ACLY2
25ACLY15
25ACLY3
25ACLY8
25ACLY9
25ACLY5
25ACLY10
25ACLY1
25ACLY11
25ACLY6
25ATALB1
25ATALB5
25ATALB3
Line #
B.
r = -0.66, p = 3.2e-05
#18
Area [mm²]
150
100
50
0
−50
0.00
0.01
0.02
0.03
% PHB/gFW
C.
#18
#18ACLY8
#18ACLY8b
#25ACLY1
#25
#18ACLY9
#18ACLY9b
#25ACLY10
Tak2
#18ACLY6
#18ACLY6b
#25ACLY11
D.
% PHB/gFW
0.04
0.03
0.02
0.01
0.00
**
18
18ACLY8
18ACLY6
18ACLY9
Line #
E.
Total PHB per plant [mg]
0.012
0.010
0.008
0.006
0.004
0.002
0.000
n.s.
18
18ACLY8
18ACLY6
18ACLY9
Line #

5.4.3. The effect of the overexpression of ACLY3KQ and *At*tDT on overall plant metabolite profile

To gain further insight into the impact of PHB production and the ancillary interventions on the metabolism of *M. polymorpha*, lines #18, 18ACLY6, 18ACLY8, and 18ACLY9 as well as WT Tak1 and Tak2 were subject to a metabolite screen. Briefly, the prevalence of a panel of 21 metabolites consisting of carboxylic acids and amino acids was quantified in plants grown from gemmae for 31 d. For a full dataset of all 21 metabolites see Figure S5.3. A PCA was used to give an insight into the structure of the entire dataset. After data scaling, a scree plot showed that the two first principal components accounted for 80 % of the variance in the data, with principal component 1 (PC1) accounting for 70.3 % and principal component 2 (PC2) accounting for 9.7 %. Principal components 1 and 2 were plotted against each other to cluster samples based on their similarity.

The results of the PCA are shown in Figure 5.5B. The samples from Tak1 and Tak2 plants cluster differently showing that the metabolite profiles in the male and female WT lines differ considerably. Given that line #18 is known to be Tak2, further comparisons against WT included only Tak2. The metabolites can clearly be used to distinguish WT plants from plants producing PHB; distinguishing between the constitutive line and the three ancillary lines is less clear due to the variation between lines #18ACLY6, 8 and 9 with line #18ACLY9 being the most clearly different to constitutive line #18.

A loading plot was constructed to determine how much each metabolite influenced PC1 and PC2 (Figure 5.5A). Loadings are computed for each variable and range from -

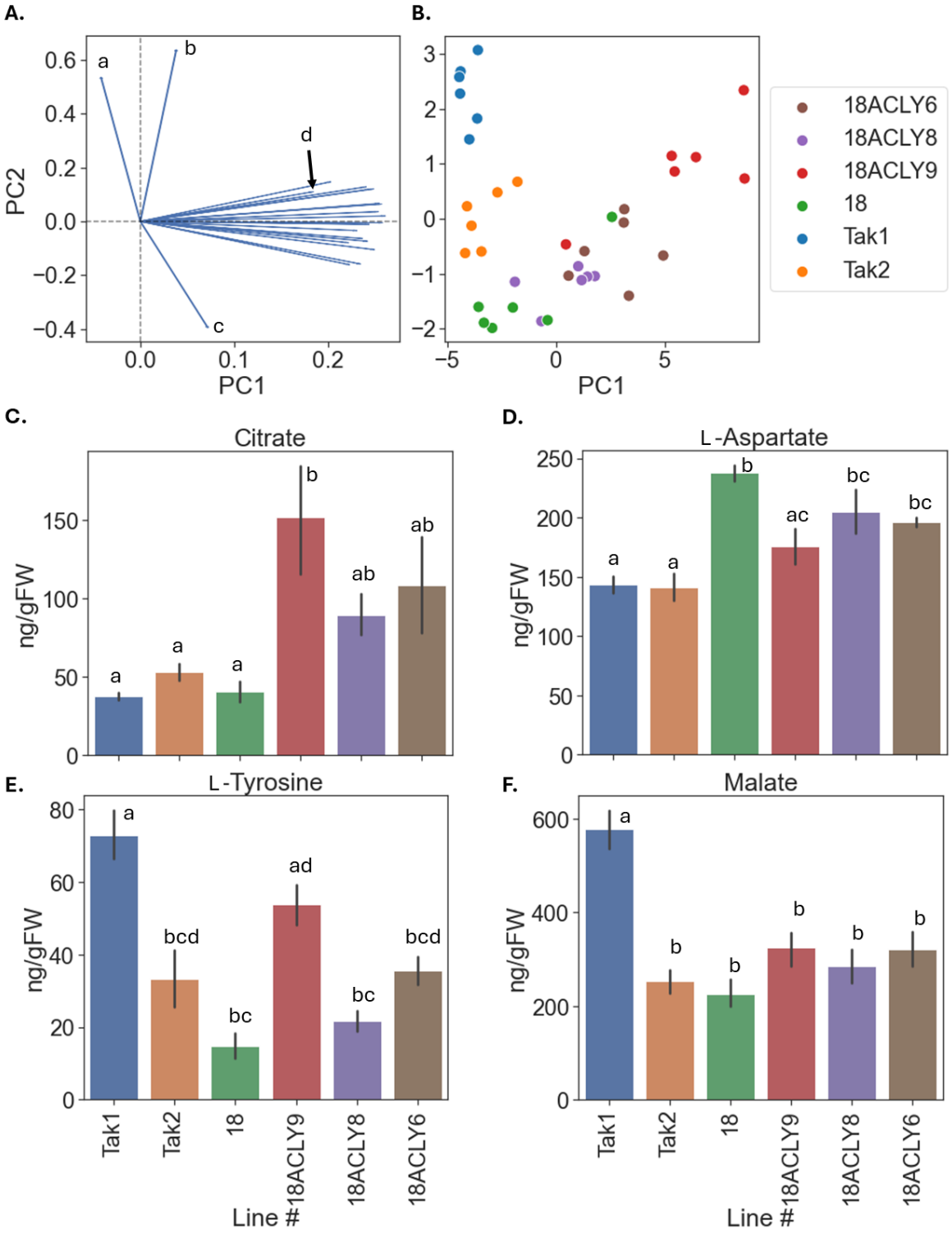
1 to 1. A high absolute value describes a variable that strongly influences the component. Conversely, values close to 0 indicate that the variable has a weak influence of the component. In this analysis, loadings ranged from approximately 0.2 to 0.6. Malate, L-tyrosine, and L-aspartate have the most effect on PC2 with loadings between approximately 0.4 and 0.6, and the remaining metabolites have the most effect on PC1 with loadings of approximately 0.2. The angles between the vectors in a loading plot can also provide information regarding how characteristics correlate with each other. When two vectors are close and the angle between them thereby small, they are likely to be correlated; this is true of the group of 18 metabolites strongly affecting PC1. If they meet each other at or close to 90 ° they are not likely to be correlated; this is true of the previously mentioned 18 metabolites and malate, L-tyrosine, and L-aspartate. Lastly, when the loading vectors diverge and form an angle close to 180 ° they are negatively correlated, such as malate and L-aspartate.

Finally, the sign of a loading value indicates whether a variable and a principal component are positively or negatively correlated. Hence PC2 is negatively correlated with L-aspartate, but positively correlated with malate and L-tyrosine. Given that the largest variation between WT and line #18 is along the PC2 axis, and that Tak2 samples have higher PC2 values, we can conclude that the biggest difference between Tak2 and constitutive PHB-producing line #18 is a relative decrease in malate and L-tyrosine but an increase in L-aspartate with the production of PHB. This is also evident in the individual plots for L-aspartate and L-tyrosine (Figure 5.5D and E).

Tak2 and the ancillary lines vary mostly along the PC1 axis (Figure 5.5A). PC1 is positively correlated with all metabolites apart from malate, L-tyrosine, and L-

aspartate, including citrate (Figure 5.5A). Indeed, levels of these metabolites are generally elevated in the ancillary lines (Figure 5.5C and Figure S5.3).

Figure 5.5 (next page): The effect of the overexpression of ACLY3KQ and *At*tDT on overall plant metabolite profile. (A) Loading plot showing the impact of each metabolite on both PC1 and PC2. a, malate; b, L-tyrosine; c, L-aspartate; d, citrate. (B) Clustering of samples in PCA based on variance accounted for by PC1 and PC2. (C-F) Contents of citrate (C), L-aspartate (D), L-tyrosine (E), and malate (F) in WT, constitutive PHB-producing line #18, and three ancillary lines. n=6 in all cases. Error bars represent SEM. In all cases the groups were shown to be significantly different from each other (ANOVA, $P < 0.05$), and a post-hoc Tukey's test used to determine differences between specific conditions. Significantly different conditions are indicated using letters, where two conditions with no overlap in letters are significantly different to each other ($P < 0.05$).


A.
B.
a
b
c
d
PC1
PC2
18ACLY6
18ACLY8
18ACLY9
18
Tak1
Tak2
C.
Citrate
D.
L-Aspartate
E.
L-Tyrosine
F.
Malate
ng/gFW
Line #
Tak1
Tak2
18
18ACLY9
18ACLY8
18ACLY6

5.5. Discussion

5.5.1. Negative effects associated with PHB production in *M. polymorpha*

It is thought that the main cause of growth reduction in plants producing cytosolic PHB is the competition between PHB and biomass-accumulating pathways for acetyl-CoA (*56, 63, 193*). Indeed, the activity of cytosolic ACL, which catalyses the production of acetyl-CoA and oxaloacetate from citrate and CoA with hydrolysis of ATP, was shown to be key for cytosolic PHB production in a model of the central metabolism of *M. polymorpha* (see Chapter 3). The competition between PHB production and biomass accumulation was strikingly evident in *M. polymorpha* lines constitutively producing PHB and there was a strong negative correlation between plant size and PHB accumulation (Figure 5.1B).

The plant lines constitutively producing PHB displayed a range of phenotypes (Figure S5.1). As well as reduced plant size, the production of PHB was also associated with the presence of large openings in the dorsal epidermis of the thallus, hypothesized to be enlarged air pores. These were visible in several lines of plants transformed with constitutive PHB, inducible PHB (but only after induction), or ancillary intervention constructs (Figures 5.1C, 5.2C, and 5.4C, respectively) but never in WT plants. The reason for these defects is unclear. Air pores allow gas exchange; it is possible, therefore, that enlarged air pores might arise as a result of carbon limitation, perhaps to allow for higher rates of photosynthesis analogous to an increased number of stomata in higher plants, shown to improve photosynthetic induction and biomass

production (*263*). It is more likely, however, that these structures are a result of reduced levels of available cytosolic acetyl-CoA, given the production of PHB. Cytosolic acetyl-CoA generated by ACL is essential for the acetylation of cell wall polysaccharides such as glucomannan, xyloglucan, and pectin, and the downregulation of ACL genes in *A. thaliana* has been shown to result in severe alteration in secondary wall structure of xylem vessels (*114*).

Finally, as well as cell wall acetylation, cytosolic acetyl-CoA is involved in the synthesis of compounds such as waxes and sterols. Overexpression of ACL in *Arabidopsis* led to a 30 % increase in wax deposition on stems, and transgenic *Arabidopsis* plants producing PHB were shown to have a 50 % reduced wax content, an effect which was rescued by ACL overexpression in this background. In addition, the analysis of cutin-specific monomers in rosette leaves revealed a 50% reduction in PHB-producing lines compared to WT, an effect again at least partly rescued by the overexpression of ACL in this background (*61*). As such both main components of the plant cuticle relies on cytosolic ACL activity and are negatively affected by PHB accumulation. This may also partly affect the abnormalities observed in *Marchantia* plants producing PHB. To determine whether PHB production affects wax and cutin biosynthesis in *Marchantia*, these components could be quantified and tests such as toluene blue staining used to quantify cuticular permeability (*61, 264*).

5.5.2. Can inducible production alleviate phenotypic defects associated with PHB production?

In the constitutive lines negative phenotypes were expected; these lines were to be

used as a baseline against which inducible PHB production and the introduction of computationally derived engineering targets could be evaluated. Inducible systems are often used in metabolic engineering to overcome deleterious or lethal effects of product accumulation or metabolic disruption. Inducibility has also been used in plant based PHB production to avoid competition between PHB production and endogenous metabolism (*46-48*). Chemically-induced promoters such as with ethanol (*70*), salicylic acid (*71*), or ecdysone (*72*) have been used to drive plastidial PHB production and improve plant phenotypes. To my knowledge, however, no inducible system has been used for cytosolic PHB production. In this work, the PULSE optogenetic system was used to drive the production of cytosolic PHB in response to red light.

The functionality of PULSE in *M. polymorpha* was established using luciferase as a reporter and is described in Chapter 4. In this chapter the production of PHB was shown to be fully inducible and minimally leaky (Figures 5.2 and 5.3). However, inducible lines grown from gemmae also had enlarged pore-like structures after induction and were 61-85 % smaller than WT Tak1 plants. In fact, they showed similar or higher levels of growth reduction to constitutive PHB lines (Figure S5.2). Accordingly, although in terms of %PHB gFW^{-1} they were more effective than constitutive lines grown in the same conditions at producing PHB, this benefit was no longer apparent once final plant size was taken into account (Figure 5.2G and H). PHB yield was reported in units of %PHB gFW^{-1} because this is the unit most prevalent in the literature. Expressed this way, differences in PHB yield can also be used to infer the preferential allocation of carbon and other cellular resources towards PHB given that

a higher value of %PHB gFW^{-1} means a larger proportion of plant mass is composed of PHB, and vice versa. As such, inducible PHB production was effective at re-allocating plant resources towards PHB synthesis, but to the detriment of other plant processes such as biomass accumulation.

Part of the problem might be to do with the induction conditions, given that three three-day exposures to red light stimulated growth (Figure S5.2). Inducing plants using three-day bursts of red light thereby sets up an increased competition during this time between PHB and biomass production. In addition, growth in monochromatic red light adversely affects photosynthesis (*257*). Shorter pulses of red light, although leading to lower expression levels (Figure 4.2E), might reduce the stimulating effect on biomass accumulation and thereby lead to higher PHB accumulation levels. In addition, it would be worth testing whether white light supplemented with a large excess of monochromatic red light might allow for higher PHB production levels whilst minimizing growth effects. When characterising the PULSE system on its own, a 2:1 ratio of red to blue light resulted in similar luciferase expression levels to monochromatic red light (Figure 4.2C). Addition of a very small amount of blue light (6 %) to a monochromatic red light growth condition has been shown to be advantageous for plant growth and yield (*257*). Finally, given the effects on phenotype associated with the PULSE system on its own (Figure 4.2H) it is hard to isolate the effects due solely to PHB production in the inducible PHB lines.

The effect of inducible PHB production on growth was more pronounced when plants were grown from gemmae that when plants were grown from thallus cuttings (Figures 5.2G-H and 5.3A-B). Yield of PHB when expressed in terms of %PHB gFW^{-1} was

increased both in gemmae and in thallus cuttings compared to constitutive PHB-producing lines grown in the same condition. When expressed in terms of total mg PHB per plant, however, only plants derived from thallus cuttings maintained an advantage compared to constitutive lines. In these two situations, the plants are induced according to the same induction scheme (Figures 5.2F and 5.3C); either the first three-day induction period is 3 d after the plating of gemmae or 3 d after the plating of thallus cuttings. It is clear, therefore, that induction is happening at very different developmental stages in the two scenarios. The germination of gemmae is dependent on the plant hormone abscisic acid which regulates dormancy and germination in response to environmental cues (*265*). Taking thallus cuttings involves the removal of the apical region of the thallus, including the meristem. This apical explant continues to propagate apically through bifurcation of the meristem, but without regeneration (*235*). It is possible that either prolonged exposure to red light or the production of PHB affects germinating gemmae and growing thalli in different ways. These effects seem the most severe in gemmae, such that inducing the production of PHB in growing thallus tissue is clearly advantageous over both induction in the same plant lines growing from gemmae, and over constitutive PHB production under the same conditions.

Even in apical thallus explants, however, PHB accumulation levels are not constant across the thallus tissue. Instead, there is a strong negative correlation (P = 0.016) between the number of induction periods in a specific segment of tissue and PHB contents in terms of %PHB gFW^{-1} (Figure S5.4A). This suggests that even in the apical thallus explants the increased rate of biomass accumulation as a result of red light is

outstripping the accumulation of PHB. This effect is probably exacerbated by the fact that older tissue is clearly less healthy than the younger tissue at the thallus apex (Figure S5.4B), even without induction. As the tissue ages, it does not maintain the ability to produce the same levels of PHB, an effect also shown in other plant species (*201, 266*). This was also visible in the heterogeneous distribution of luciferase expression when driven by PULSE (Figure 4.2A).

To conclude the PULSE system can be used to drive PHB production in *Marchantia* in response to red light, and if done in thallus cuttings this results in higher yields than constitutive lines grown in the same conditions. When compared to constitutive PHB-producing plant lines grown in normal growth conditions, however, yields are inferior (Figure 5.1E and 5.3A). This is probably due to a combination of the negative effects of the PULSE system itself on plant growth, as well as the growth-promoting effects of monochromatic red light.

5.5.3. The effects of computationally derived targets on the ability of *Marchantia* to support cytosolic PHB production

The introduction of *At*tDT and ACLY3KQ into a PHB-producing plant background was based on a model used to find the optimal path to cytosolic PHB production whilst allowing growth. In the model, the flux to PHB was constrained to 10 % of the flux to biomass, and this required a 12.7-fold increase in ACL activity. Unfortunately, this work did not extend to a quantification of ACL activity in the ancillary lines, but previous ACL overexpression studies showed that a 2.7-fold increase in ACL activity was able to fully rescue phenotypes associated with PHB production in *Arabidopsis*, but did not

significantly increase PHB yield in terms of %PHB gFW^{-1} *(61)*. In the ancillary lines presented here, the effects were the opposite of this; growth was severely negatively affected both in comparison to the parent constitutive line and to WT plants (Figure 5.4B), but when expressed in terms of %PHB gFW^{-1}, PHB yields in lines carrying ancillary interventions were up to 4.7-fold higher than in parent line #18 (line #18ACLY9, Figure 5.4D). Due to a combination of these two effects, and in a similar way to the inducible lines, the apparent benefit to PHB production was ablated when considering final plant size (Figure 5.4E).

Given the combinatorial ways in which *At*tDT and ACLY3KQ might be affecting both growth and PHB production, it is difficult to identify the cause of the opposite result to published work regarding the effect of ACL overexpression on plant based PHB production obtained here. It becomes even more challenging given the gaps still present in the characterisation of the ancillary lines, including confirmation of protein expression and activity. For instance, it is possible that ACLY3KQ is inactive in *Marchantia*, and that the increase in %PHB gFW^{-1} is purely the result of increased citrate efflux from the vacuole due to *At*tDT overexpression (although this becomes unlikely when considering results of the metabolite screen, see 5.5.4). Alternatively, perhaps overexpression of ACLY3KQ is encouraging the production of PHB, but the activity of *At*tDT is having negative effects on growth by disrupting a diurnal vacuolar carboxylic acid storage pattern. These targets need to be expressed individually as well as in combination to determine how they are affecting final plant size as well as PHB production in *Marchantia*.

Nevertheless, even though the cause cannot be precisely pinpointed, it is clear that

the computationally derived targets are affecting PHB production. As already explained, expressing PHB yield in terms of %PHB gFW^{-1} can be used to infer the preferential allocation of carbon and other cellular resources towards PHB given that a higher value of %PHB gFW^{-1} means a larger proportion of plant mass is composed of PHB, and vice versa. We can therefore conclude that the ancillary targets resulted in greater allocation of plant resources towards PHB synthesis when compared to constitutive PHB production on its own. Perhaps the high expression of the PHB synthesis enzymes means that any acetyl-CoA produced by ACLY3KQ is used up by β-ketothiolase before it can be used by endogenous reactions. Alternatively, perhaps the high expression of ACLY3KQ is causing a decrease in intracellular ATP concentrations which is adversely affecting endogenous metabolism more than PHB biosynthesis.

In a similar way, if the overexpression of ACLY3KQ has been effective at relieving competition for acetyl-CoA, the competition between PHB and biomass accumulation may have been shifted to NADPH, the co-factor required for acetoacetyl-CoA reductase activity. The computational modelling of cytosolic PHB production in *Marchantia* showed that there was a perfect positive correlation between flux to cytosolic PHB accumulation and the GAPN reaction which represents a variant of glycolysis that conserves energy as NADPH rather than as ATP (section 3.5.1). In addition, the analysis of a stoichiometric model of *S. cerevisiae* expressing the three enzymes required for PHB synthesis during slow growth suggested that the theoretical yield of PHB could be improved by including a combination of ACL and a transhydrogenase reaction to catalyse the conversion of NADH and $NADP^+$ to NAD^+ and NADPH, respectively (*194*). Hence NADPH seems important to ensuring simultaneous

growth and PHB production and should be considered as a potential additional engineering target.

5.5.4. The effects of PHB production on the metabolism of *M. polymorpha*

The metabolite profile of a plant can provide a snapshot insight into its metabolism. In this work, a panel of 21 metabolites was quantified in WT, constitutive PHB, and ancillary *Marchantia* lines. Using PCA, I showed that the metabolite profile can be used to differentiate between WT Tak2 plants and constitutive PHB-producing line #18, and that this difference is mostly accounted for by increased L-aspartate and decreased L-tyrosine in the PHB-producing line (Figure 5.5). Given the complexity of metabolism it is difficult to pinpoint the exact reasons for differences between conditions of the concentration of specific metabolites. Nonetheless, the difference in abundance of L-aspartate and L-tyrosine between WT and PHB-producing scenarios might be explained in terms of the PHB biosynthetic pathway.

L-tyrosine is downstream of the shikimate pathway, which generates chorismate from phosphoenolpyruvate and D-erythrose 4-phosphate and via shikimate. Chorismate is converted to prephenate by chorismate mutase, which is the precursor to both L-tyrosine and L-phenylalanine. To convert prephenate into L-tyrosine, plants use mainly the arogenate pathway, in which the final step is arogenate dehydrogenase (*267, 268*). Hence the production of L-tyrosine is highly dependent on shikimate; the production of shikimate requires the NADPH-dependent shikimate dehydrogenase to catalyse the reduction of 3-dehydroshikimate (*269*). In fact, many attempts at engineering

microorganisms for L-tyrosine production have involved increased flux through the shikimate pathway, and the highest yields been obtained when sufficient NADPH supply was also engineered (*270*). As described the production of PHB also requires NADPH (Figure 1.1A). It is possible, therefore, that the production of PHB reduces the availability of NADPH to shikimate dehydrogenase, thereby reducing the production of chorismate and downstream L-tyrosine. If true, this would provide further evidence that NADPH provision should be considered an additional engineering target. This does not explain, however, why L-phenylalanine and L-tryptophan contents are not also reduced in constitutive PHB line #18 (see Figure S5.3) given that they are also derived from chorismate and the shikimate pathway.

The levels of L-aspartate are significantly elevated in the constitutive PHB-producing line compared to WT. L-aspartate is synthesized mainly from oxaloacetate (OAA) by aspartate aminotransferase. Given that ACLY3KQ produces OAA as well as acetyl-CoA, the increase in L-aspartate contents in the ancillary lines compared to WT could be due to increased provision of OAA to aspartate aminotransferase. Why this effect would also be visible in a constitutive PHB-producing line is less clear, but may be related to the fact that ACL is post-translationally regulated by sink strength (*61*). As such driving high levels of PHB production causes an increase in endogenous ACL activity, thereby resulting in higher cytosolic OAA levels which can be converted to L-aspartate.

To my knowledge only one other study of the effect of PHB production in plants on the plant metabolome exists, and that was concerned with plastidial production (*52*). In that study, increased PHB accumulation in *Arabidopsis* lines correlated with

decreased levels of fumarate and isocitrate, and an increase in osmoprotectants such as L-proline and mannitol as well as sugars such as glucose, fructose and sucrose. Of these, only L-proline was quantified in this work, levels of which were indeed increased in all PHB-producing lines (Figure S5.3).

Conclusions made about the metabolite profile associated with constitutive PHB production need to be tempered by the fact that only one line has been analysed. The intraspecific variation in metabolite profiles shown between ancillary lines (and even between different samples from the same line) indicates that more constitutive lines should be analysed before any firm conclusions made. Such an approach would also allow for an exploration of the correlation between accumulated PHB levels and key metabolite concentrations.

Of the 21 metabolites analysed, there were 18 which were highly positively correlated with PC1. PC1 accounted for most of the variation between WT and constitutive-PHB producing line #18 and the three ancillary lines tested (Figure 5.5B). These 18 metabolites were all amino acids, apart from citrate and glycyl-L-proline, and were for the most part elevated in the ancillary lines (Figure S5.3). In general, the abundance of each amino acid followed the same pattern as the PHB contents in the ancillary lines, i.e. highest in #18ACLY9, followed by #18ACLY6 and lowest in #18ACLY8. Plants are known to accumulate amino acids in response to stress (*271*). In addition, plants use the vacuolar lytic degradation of proteins to release free amino acids during starvation, which the plant can use instead of sugars for respiratory energy production (*272*). As such the elevated levels of amino acids in the ancillary lines compared to constitutive PHB and WT plants may be due to adverse effects on the plants due to the

accumulation of PHB which effectively acts as an extra carbon sink coupled with metabolic disruption due to the ancillary targets themselves.

Interestingly, citrate levels are elevated in the ancillary lines compared to the other plant backgrounds. In the only previous overexpression study of *At*tDT, citrate levels decreased in the *Arabidopsis* tissue. This was postulated to be due to export of citrate from the vacuole and subsequent catabolism of citrate in the cytosol (*181*). Again, we are limited in the conclusions that we can draw based on the lack of expression and activity data. The increase in citrate contents may be due to a combinatorial effect of overexpression of ACLY3KQ alongside *At*tDT. As with the effects on PHB accumulation and plant growth, the effects on metabolism could be more clearly established if ACLY3KQ and *At*tDT were expressed individually, and if they were also overexpressed in a WT background.

To conclude, the unexpected and marked consequences associated with PHB accumulation as well as the expression of the ancillary constructs on metabolite profiles provides an illustration of the interconnectedness and complexity of metabolism.

6. Building synthetic biomolecular condensates in plants

6.1. Introduction

Compartmentalisation is used throughout the tree of life to regulate and control metabolism. The canonical example of an intracellular compartment is the membrane-bound organelle, but cells also use membraneless organelles such as biomolecular condensates to compartmentalise metabolism (*75, 92*). Although a large body of research has emerged over the last 10 years regarding these structures, questions still remain as to their biophysical properties. For example, not much is known about how the diffusion barrier presented by these structures might influence the entry and exit of metabolites. In addition, debate exists as to whether biomolecular condensates form exclusively through LLPS, and how to experimentally determine this *in vivo*.

What has become clear, however, is that by compartmentalising metabolism using biomolecular condensates, cells can both control metabolic branch points and selectively provide kinetic benefits to the recruited enzymes (*77*). For instance, concentrating enzymes and their substrates together can increase enzyme efficiency via mass action (*78*). On top of this, and in the case of multiple enzymes forming part of a pathway, clustering these enzymes together can enable probabilistic substrate channelling (*79*). This phenomenon occurs because when a product diffuses away

from the active site of one enzyme in the cluster, the probability of encountering the next enzyme in the pathway is much higher than the metabolite diffusing into the bulk aqueous solvent of the cell.

Building synthetic biomolecular condensates and 'functionalising' them by the recruitment of enzymes has been explored as a way of controlling metabolic branchpoints and increasing flux through a pathway, both important goals in metabolic engineering. Certain classes of molecules possess a propensity for the formation of biomolecular condensates due to their ability to form multivalent interactions with each other. For example, many protein IDRs have been shown to form biomolecular condensates both *in vitro* and *in vivo (103, 106-108, 182)*. These have therefore been used to generate synthetic biomolecular condensates in a range of organisms. For example, the OptoDroplet system uses light to trigger IDR-based condensate formation and has been used to assemble enzymes of the deoxyviolacein pathway into a biomolecular condensate using light in yeast. By doing this the yield increased 6-fold because it favoured flux through the deoxyviolacein pathway by offering control over a metabolic branch point (*6*). In addition, recruiting β-galactosidase to a condensate built from artificial IDPs was shown to provide a statistically significant increase in the catalytic constant or turnover efficiency (k_{cat}) of the enzyme in *E. coli* and human embryonic kidney cells (*107*).

In this chapter I will describe how synthetic biomolecular condensates were designed and built *in planta*. The condensates were functionalised with enzymes in order to determine their usefulness for plant metabolic engineering. For example, the low yield of PHB usually achieved in plants might be overcome by increasing flux through the

PHB production pathway. Similarly, controlling the metabolic flux point at acetyl-CoA could be used to increase flux through PHB synthesis.

6.2. The design of a construct for building synthetic biomolecular condensates in plants

A construct for the creation of synthetic biomolecular condensates in plants was designed based on published work (*109*). The system is composed of a scaffold and a client. The scaffold was built from two copies of RGG, the IDR from the *C. elegans* LAF-1 protein which is involved in P-granule formation. The role of the RGG IDR in driving phase-separation during P-granule formation is well-documented (*273*), and this domain has previously been used to generate synthetic biomolecular condensates in yeast and human cells (*104-106, 109*). The monomeric green/yellow GFP-derived fluorescent protein (FP) mClover3 was included for visualization (*173*), and one of a cognate pair of synthetic coiled-coil peptides was included for client recruitment (SYNZIP1) (*183*). The client comprised the monomeric red FP mCherry (*274*), and the other element of the cognate pair of synthetic coiled-coils (SYNZIP2). Monomeric FPs are recommended in this type of experiment as multimerising activity might affect condensate formation (*275*). Furthermore, mClover3 and mCherry are expected to have very limited spectral overlap making them ideal for co-localisation experiments (*276*). To aid in purification using Ni-affinity purification, both the scaffold and client also included a 10×His tag on either the C or the N terminus, respectively. The client and scaffold system was designed as part of a polycistronic construct separated by the viral self-cleaving InteinF2A peptide (see section 5.2.1). After translation, the

scaffold and client should form two separate fusion proteins; the scaffold protein should assemble into biomolecular condensates via the multiple and weak interactions between RGG domains, and mCherry should be recruited to the condensate via the SYNZIP pair dimerisation. The scaffold-client system is illustrated in Figure 6.1A.

Two versions of this scaffold-client system were designed, the difference being the presence or absence of a nuclear localisation signal (NLS) on the C-terminus of mClover3. The version of the construct including the NLS was expected to form condensates in the nucleus, whereas the version with no NLS was expected to form cytosolic biomolecular condensates. In addition, both the nuclear and cytosolic versions of the construct were modified for the removal of the RGG domains as a control. The versions without the RGG domains were not expected to form biomolecular condensates in either location.

The different versions of the scaffold-client system were cloned into both the pK7WG2 (*174*) and p*Mp*GWB103 (*42*) Gateway expression vectors. pK7WG2 places the construct downstream of the constitutive p35S promoter and was used for transient expression in *N. benthamiana*, whereas p*Mp*GWB103 is specifically designed for overexpression in *M. polymorpha* and places the coding sequence downstream of the strong endogenous p*Mp*EF1α promoter and carries hygromycin resistance.

6.3. Characterisation of synthetic biomolecular condensates *in planta*

I had initially aimed to characterise the construction of biomolecular condensates in

M. polymorpha. However, despite two different sporeling transformation attempts in which transformants were recovered and confirmed by genotyping, no biomolecular condensates were observed either in the nucleus or cytosol of any of the plants screened. Possible reasons for this will be discussed in section 6.5.

Therefore, to characterise the feasibility of building synthetic biomolecular condensates *in planta* I switched to the use of transient expression of the construct in *N. benthamiana*. This approach also has the benefit of allowing for the rapid and combinatorial testing of different constructs. The nuclear and cytosolic versions of the construct were agroinfiltrated into the underside of mature *N. benthamiana* leaves and expression was monitored over 7 d. Biomolecular condensates were visible as soon as 2 days post-infiltration (dpi). In both cases expression levels increased until day 3 or 5 post-infiltration and then decreased, such that barely any signal was visible after day 7 (Figure S6.1A). The exception to this was in some cells transformed with the cytosolic version of the construct where a small number of very large biomolecular condensates were occasionally still seen 7 dpi (Figure S6.1C). All subsequent characterisation was carried out on leaf tissue at 5 dpi.

Punctate structures were observed in both nuclei and cytosol of transiently transformed *N. benthamiana* leaves when infiltrated with the nuclear or cytosolic version of the construct, respectively (Figure 6.1B, C). Nuclear localisation was confirmed using DAPI staining (Figure S6.1B). There was good signal overlap between the mClover3 and mCherry channels indicating that in many cases these two proteins co-localised (Figures 6.1C), although this should be verified using a statistical co-localisation test such as the Manders overlap coefficient (*277*). The fact that mCherry

signal was detected in the nucleus suggests that client recruitment via SYNZIP dimerisation may be occurring in the cytosol before recognition of the NLS on the scaffold and subsequent trafficking into the nucleus. Given that proteins of the correct size for the client and scaffold part of the construct were detected in the leaf crude lysate (Figure 6.1F), I concluded that the polycistronic construct was cleaving as expected, and that the interaction between SYNZIP1 and SYNZIP2 was effectively recruiting the client into the structure formed by the scaffold. In order to confirm that the punctate structures were forming due to interactions between RGG domains, versions of the nuclear and cytosolic construct without these domains were constructed, infiltrated into *N. benthamiana* leaves, and expression over 7 d was monitored as described. In both the cytosol and the nucleus diffuse signal was observed and increased until 3 dpi. No punctate structures were observed in plants infiltrated with these constructs. A single large and diffuse structure was observed in some nuclei, posited to be the nucleolus (Figure 6.1B).

The abilities of the cytosolic and nuclear versions of the construct for biomolecular condensate formation were compared by calculating the number of biomolecular condensates in each cell as well as the volume of each of these biomolecular condensates (Figure 6.1D). There was a large and significant ($P < 0.001$) difference between the two locations for both metrics; on average, there were 7.7 biomolecular condensates per nucleus, but 41.9 in the cytosol. In some cases, in the cytosol, there were up to 175 biomolecular condensates in a single cell. The volume of the biomolecular condensates was also significantly larger in the cytosol than in the nucleus. The mean volume of a nuclear biomolecular condensate was 0.47 μm^3

whereas the mean cytosolic biomolecular condensate volume was 5.7 μm^3. The largest cytosolic biomolecular condensate measured was 495 μm^3. For context, this is approximately the same size as the nucleus in *Triticum aestivum* (424 μm^3) and roughly three and a half times the size of the nucleus in *N. tabacum* (149 μm^3) (*278*). In plant cells the volume of the cytosol is expected to be approximately 30-fold larger than the volume of the nucleus (*279*); as such it is not unexpected that it can accommodate both more and larger biomolecular condensates. In addition, for biomolecular condensates to be formed in the nucleus the scaffold protein has to be trafficked into the nucleus which introduces an extra obstacle to biomolecular condensate formation.

FRAP has been used extensively in the biomolecular condensate field to determine whether a condensate is 'liquid-like' and thereby likely to have formed through LLPS. Recent work has disputed the ability of FRAP to determine anything other than the mobility of the molecules into and inside condensates (*92, 280*), however, and this will be discussed more in section 6.5. As such in this work I used FRAP to determine the molecular mobility of the scaffold protein in the condensate. A recovery curve can yield a percentage recovery value and a rate of recovery; percentage recovery provides information regarding the proportion of molecules in the bleached area which are mobile, whilst the recovery rate constant can provide information regarding the speed at which molecules move in and out of the bleached area (*281*). The recovery rate is related to the diffusion coefficient, which can be calculated from the recovery curve using the equation outlined in section 2.6.4. The diffusion coefficient can provide information regarding how rapidly diffusion is occurring by providing a measure of the

mean squared displacement per unit time of the diffusing species (*282*).

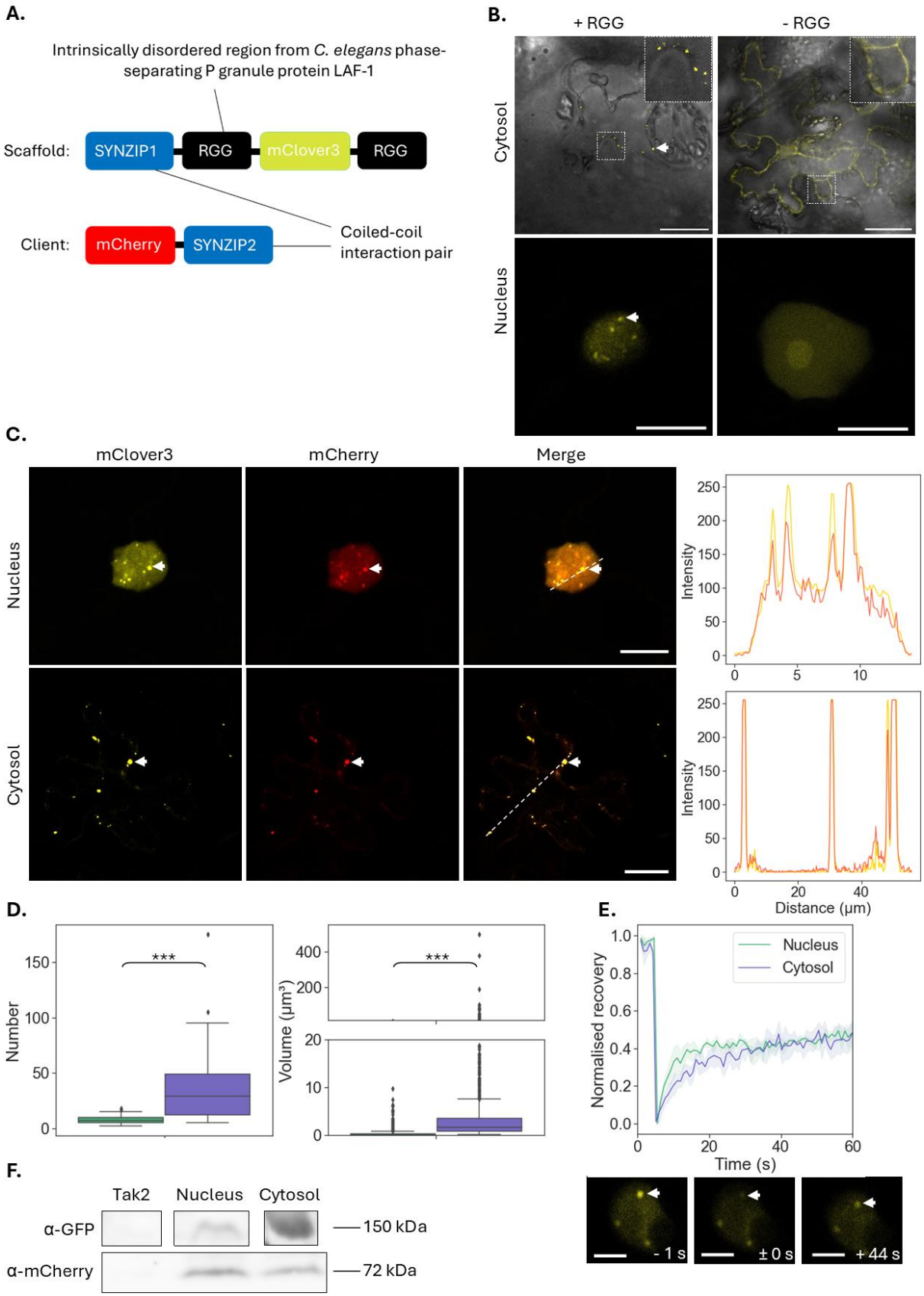
Given the size of most of the biomolecular condensates only a full-FRAP analysis could be undertaken, i.e. the entire biomolecular condensate was bleached, and the recovery of the entire biomolecular condensate was monitored. The recovery rate of a full FRAP can provide information regarding the exchange rate between the dilute and the dense phases, but is also affected by other parameters such as size of photobleached area, size of the droplet, and mobility within the droplet (*280*), so should be interpreted with caution. Preliminary work using half-FRAP (where only part of the structure is bleached) of the very large biomolecular condensates seen in the cytosol was attempted but not completed in time to be included here.

The recovery rate (i.e. the slope of the recovery curve) was slightly higher in the nucleus than in the cytosol resulting in average recovery times (i.e. the time taken for fluorescence recovery to reach steady state) of 10.3 and 27.3 s, respectively. In both subcellular locations the recovery percentage was roughly 45 % (Figure 6.1E). As such the biomolecular condensate is showing intermediate mobility and is partly formed of an immobile fraction. The fact that the recovery percentage is similar in the nucleus and the cytosol suggests that this is an inherent property of the biomolecular condensate itself which isn't affected by subcellular location. When expressed in *E. coli*, a similar RGG-based biomolecular condensate showed 40 % recovery in a full-FRAP experiment (*105*). Based on the analysis of recovery curves from six and seven nuclear and cytosolic biomolecular condensates, respectively, diffusion coefficients of 0.13 and 0.059 $\mu m^2\ s^{-1}$ were calculated. This is in line with a previous analysis of RGG-based biomolecular condensates *in vitro* which calculated a diffusion coefficient

of 0.085 μm^2 s^{-1}, although that was a half-FRAP experiment so is arguably quantifying a different type of mobility (*191*).

To conclude, RGG-based biomolecular condensates were successfully generated in the nucleus and cytosol of plant cells. These biomolecular condensates were shown to recruit a FP as a client and behaved similarly to RGG-based synthetic condensates in other systems. Given the formation of more and larger condensates in the cytosol than in the nucleus, and the greater number of potential applications for cytosolic engineering, only the cytosolic version of the construct was considered going forward.

Figure 6.1 (next page): Characterisation of synthetic biomolecular condensates *in planta*. (A) An illustration of the scaffold-client system designed and used to generate synthetic biomolecular condensates in plants. (B) The effect of removing RGG domains on the presence of biomolecular condensates in the nucleus and cytosol of agroinfiltrated *N. benthamiana* leaves. Biomolecular condensates indicated by white arrows. Scale bar is 50 µm (cytosol) and 10 µm (nucleus). (C) Left: Punctate fluorescence indicates likely biomolecular condensate formation in both nuclei and cytosol of agroinfiltrated *N. benthamiana* leaves, as indicated using a white arrow. Right: Signal intensity of both mClover3 (yellow) and mCherry (red) plotted across a transect illustrated in merged image. Scale bar is 10 µm (nucleus) and 20 µm (cytosol). (D) A comparison of the number per cell (left) and volume (right) of cytosolic (blue) and nuclear (green) biomolecular condensates. The distribution is plotted using a box and whisker plot, where the box shows the mean and the interquartile range. The whiskers extend to values above or below 1.5 × the interquartile range. Points either above or below this value are shown as diamonds. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U). n = 21-54 (number) or 400-880 (volume). (E) Normalised FRAP curves for nuclear and cytosolic biomolecular condensates (top), and example of biomolecular condensate bleaching in a nucleus (bottom). n = 4-5. Arrow indicates bleached biomolecular condensate. Pre (-) or post (+) bleaching time indicated. Scale bar is 10 µm and shaded area represents SEM. (F) Western blot for the detection of mClover3 (α-GFP) and mCherry in crude leaf extract of agroinfiltrated *N. benthamiana* leaves 5 days after infiltration.


A.
Intrinsically disordered region from C. elegans phase-separating P granule protein LAF-1
Scaffold:
SYNZIP1
RGG
mClover3
RGG
Client:
mCherry
SYNZIP2
Coiled-coil interaction pair
B.
+ RGG
- RGG
Cytosol
Nucleus
C.
mClover3
mCherry
Merge
Nucleus
Cytosol
Intensity
Distance (µm)
D.
Number
Volume (µm³)
***
E.
Normalised recovery
Time (s)
Nucleus
Cytosol
- 1 s
± 0 s
+ 44 s
F.
Tak2
Nucleus
Cytosol
α-GFP
α-mCherry
150 kDa
72 kDa

6.4. The targeting of enzymes to synthetic biomolecular condensates *in planta*

I wanted to build synthetic biomolecular condensates in plants because of their potential as a tool for plant metabolic engineering. To do this I modified the scaffold-client system already described by altering the client portion. The mCherry previously used for visualisation of client recruitment was either replaced or supplemented with one or several different enzymes. Using this system, following co-translational cleavage it was expected that the scaffold fusion protein would assemble into biomolecular condensates and the dimerisation of the coiled-coil pair would recruit one or several enzymes to the biomolecular condensates.

6.4.1. Citramalate synthase

Citramalate synthase from *Malus domestica* (*Md*CMS) was used to determine whether enzyme activity could be supported inside the RGG-based biomolecular condensates. In addition, *Md*CMS catalyses the condensation of acetyl-CoA and pyruvate to (S)-citramalate (*184*) making it a good test for the permeability of the biomolecular condensates for acetyl-CoA, and thereby whether they might support PHB synthesis. Finally, citramalate is not expected to be present in WT *N. benthamiana*, facilitating its quantification in agroinfiltrated tissue.

*Md*CMS is localised to the plastid (*184*). To construct a version of *Md*CMS that would stay in the cytosol to be targeted to the condensates, I performed an alignment analysis using EMBOSS Stretcher from EMBL-EBI (*283*) with the closely related

methylthioalkylmalate synthase 3 (MAM3) from *A. thaliana* which is known to be chloroplast localised (Figure S6.2). MAM3 has a long unstructured region on its N-terminus which is predicted to encode a chloroplast transit peptide (see UniProt accession Q9FN52). In the predicted structure for *Md*CMS, the N-terminus is also unstructured when analysed using AlphaFold (*284*) and so I hypothesized that *Md*CMS also had an N-terminal chloroplast transit peptide. TargetP-2.0 (*285*) was used to predict the presence of a N-terminal chloroplast transit peptide with high probability, including a putative cleavage site, in the *Md*CMS protein sequence. A version of *Md*CMS without the putative target peptide was used to create a cytosolically-localised *Md*CMS, and this was incorporated into the scaffold-client system as shown in Figure 6.2A.

The scaffold-client system encoding *Md*CMS with or without the putative chloroplast transit peptide was transiently transformed into *N. benthamiana* leaves. Given that most expression was previously shown to occur between 3 and 5 dpi (Figure S6.1A), samples were taken and examined on these days. As expected, cytosolic biomolecular condensates were visible in leaves both 3 and 5 dpi, and mCherry was co-localised with mClover3 signal indicating that the mCherry-*Md*CMS fusion was being recruited to the synthetic condensates (Figure 6.2C). In samples infiltrated with the version of the construct containing the *Md*CMS with a chloroplast transit peptide, mCherry signal was detected both in the cytosolic biomolecular condensates and in the chloroplasts. When the putative targeting signal was removed from *Md*CMS no mCherry signal was observed in the chloroplasts showing the predicted sequence to be a true chloroplast targeting signal. As above, removing the RGG domains resulted in no biomolecular

condensates being formed and the mClover3 signal to instead be diffusely present throughout the cytosol (Figure 6.2D).

Samples were taken from *N. benthamiana* leaves 3 and 5 dpi, and citramalate extracted and quantified as outlined in 2.8.2 (Figure 6.2B). Surprisingly, a small amount of citramalate was detected in un-infiltrated WT tissue, even though *N. benthamiana* leaves are not known to produce citramalate. Citramalate was also detected in WT *N. benthamiana* leaf tissue in previous work undertaken by András Sándor (*172*). However, both versions of *Md*CMS tested resulted in higher ($P < 0.05$) levels of citramalate than WT showing that the enzyme was active. The version of *Md*CMS with the chloroplast transit peptide resulted in much higher levels of citramalate both 3 and 5 d after infiltration than the cytosolic version. This could reflect a detrimental effect on enzyme activity by the removal of the N-terminal unstructured transit peptide but could also reflect access to higher levels of acetyl-CoA and pyruvate due to the role of the chloroplast in fatty acid synthesis. It could also illustrate a pH effect on the activity of the enzyme, given that the chloroplast stroma is more alkaline than the cytosol (*286*) and the *Md*CMS enzyme may be more active at a higher pH.

By comparing the versions of the construct with or without the inclusion of RGG domains, we can start to understand the effects of recruitment to the biomolecular condensate on *Md*CMS activity. On day 5 after infiltration there was a 3.52-fold larger amount ($P = 0.0079$) of citramalate in leaves infiltrated with the no-RGG version of the plastidial *Md*CMS than in leaves infiltrated with the RGG version. This would indicate that sequestering some of the *Md*CMS in cytosolic biomolecular condensates results

in lower amounts of citramalate being produced. However, it is unclear whether this reflects the lower substrate availability in the cytosol or whether recruitment of the enzyme to the biomolecular condensate has an effect on enzyme activity and/or substrate availability. In the case of the cytosolic *Md*CMS, there is no large difference between the citramalate contents of leaves infiltrated with either scaffold version. As such it seems there is no obvious penalty incurred on *Md*CMS activity when it is recruited into the synthetic biomolecular condensates compared to its activity in the cytosol. It would thereby follow that acetyl-CoA, H_2O, and pyruvate are able to diffuse into the condensate. This conclusion should be made with caution, however, given the lack of statistical significance between citramalate contents in most cytosolic *Md*CMS conditions and WT tissue.

To conclude, using the scaffold-client system I was able to target an enzyme to synthetic biomolecular condensate in plants. The activity of the enzyme does not seem to be adversely affected by recruitment to the condensate, although more experiments should be done to ascertain whether this is a consistently statistically significant effect.

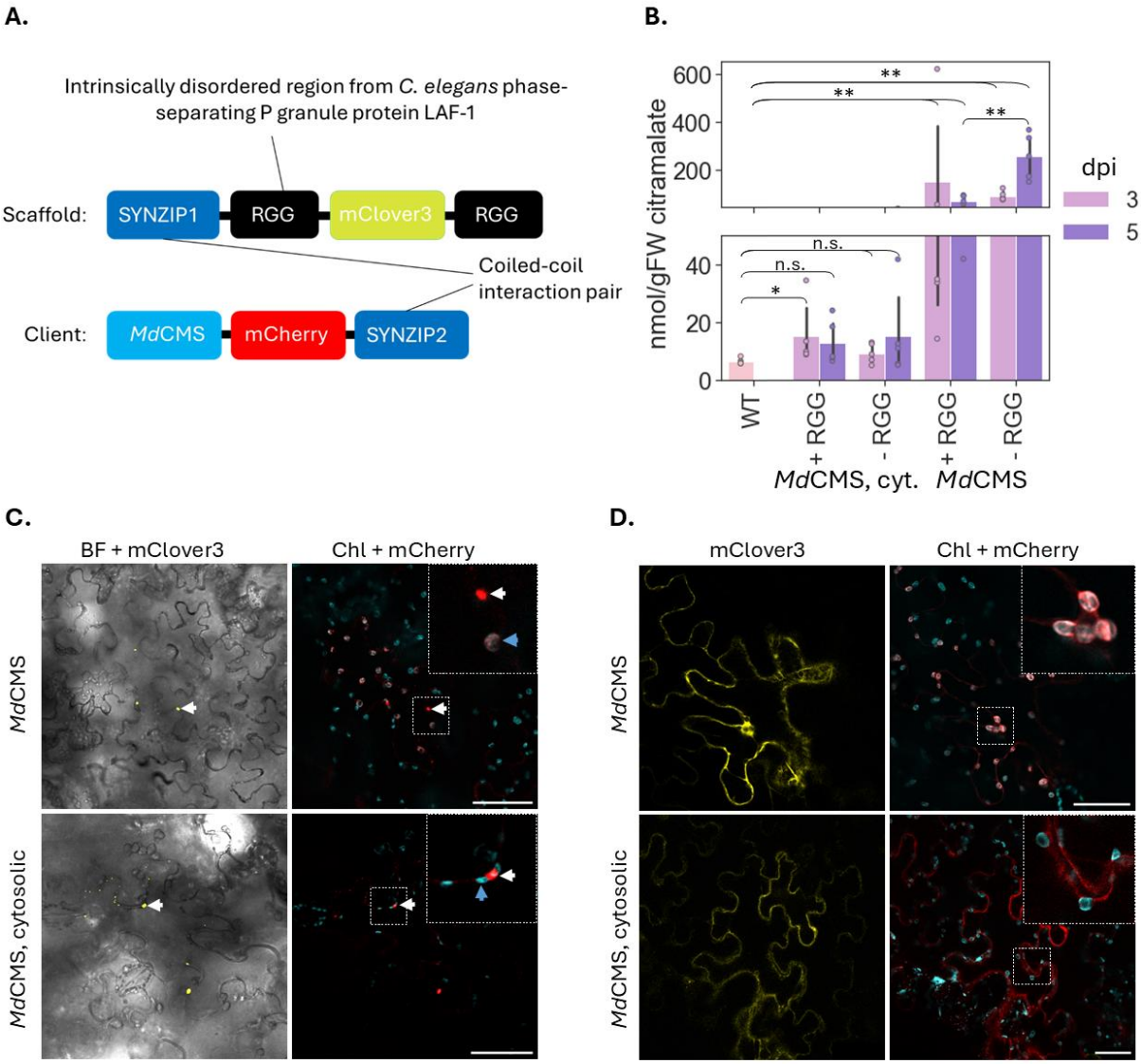


Figure 6.2: The localisation and activity of *Md*CMS in synthetic biomolecular condensates in plants. (A) An illustration of the scaffold-client system designed and used to generate synthetic biomolecular condensates to which *Md*CMS gets recruited. (B) Citramalate contents in leaves after no infiltration (WT) or infiltration with constructs for the creation of biomolecular constructs and the recruitment of *Md*CMS with or without a chloroplast targeting peptide. Both versions of *Md*CMS were also tested together with the RGG-containing (+ RGG) or no-RGG-containing (- RGG) versions of the scaffold. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U). n = 5. (C) Co-localisation of mClover3 and mCherry-fused *Md*CMS with or without a chloroplast targeting peptide. White arrow indicates a condensate, and blue arrow indicates a chloroplast. Scale bar is 100 µm. (D) The removal of the RGG domain ablates biomolecular condensate formation and results in diffuse mClover3 cytosolic signal. Scale bar is 100 µm.

6.4.2. The PHB pathway enzymes

I wanted to establish whether multiple enzymes from the same pathway could be localized to a synthetic biomolecular condensate, whether they would retain their activity, and whether organising the enzymes in this way might result in higher levels of PHB. To do this I repeated the approach used for *Md*CMS but this time I took advantage of the ability for combinatorial co-infiltration of *Agrobacterium* strains carrying different constructs. Rather than being part of the same polycistronic constrict, the scaffold and clients were designed as separate constructs. The scaffold was designed as previously described and cloned into pK7WG2. In this case, however, the three different PHB-producing enzymes were each fused to SYNZIP2 on either the N- (β-ketothiolase) or the C-terminus (acetoacetyl-CoA reductase and PHB synthase). The three fusion proteins were separated by the self-cleaving viral peptide InteinF2A as previously described. This construct, encoding the three different clients, was cloned into pK7WG2. To build biomolecular condensates and recruit the three PHB-producing enzymes, I mixed equal parts of OD_{600}-adjusted *Agrobacterium* culture containing the scaffold or the client construct. The scaffold-client system modified for PHB production is illustrated in Figure 6.3A.

It was necessary to show that all three SYNZIP2-tagged PHB-producing enzymes could be recruited to the biomolecular condensate. To do this the client-containing construct was modified to create three versions where in each case a different enzyme was fused to mCherry. This allowed the ability of the tagged enzyme to be recruited to the biomolecular condensate in the presence of the other two enzymes to be tested whilst avoiding the need for simultaneous imaging of four different fluorophores. This

method has been previously used in other systems to demonstrate the localisation of several different enzymes to condensates inside cells (*287*). All three enzymes could be successfully recruited to the biomolecular condensate in the presence of the other two (Figure 6.3C). In addition, and as previously shown, removing the RGG domains on the scaffold prevented biomolecular condensate formation and resulted in diffuse cytosolic signal from both mClover3 and mCherry (Figure 6.3D).

In order to determine whether the enzymes were functional inside the biomolecular condensates, I co-infiltrated the scaffold- and client- (without mCherry) containing *Agrobacterium* cultures into *N. benthamiana* leaves and extracted and quantified PHB. No PHB was detected in leaves 3 or 5 dpi. In order to determine whether quantifiable levels of PHB were achievable, I re-infiltrated the same leaves on day 5, and left the plants for an additional 8 d such that leaves were theoretically producing PHB for a total of 13 d. Using this method quantifiable levels of PHB were detected (Figure 6.3B). PHB levels were compared between the RGG- and no-RGG-containing scaffold conditions. There was a small (1.55-fold, $P = 0.15$) increase in PHB yield obtained when the PHB enzymes were localised to the biomolecular condensates.

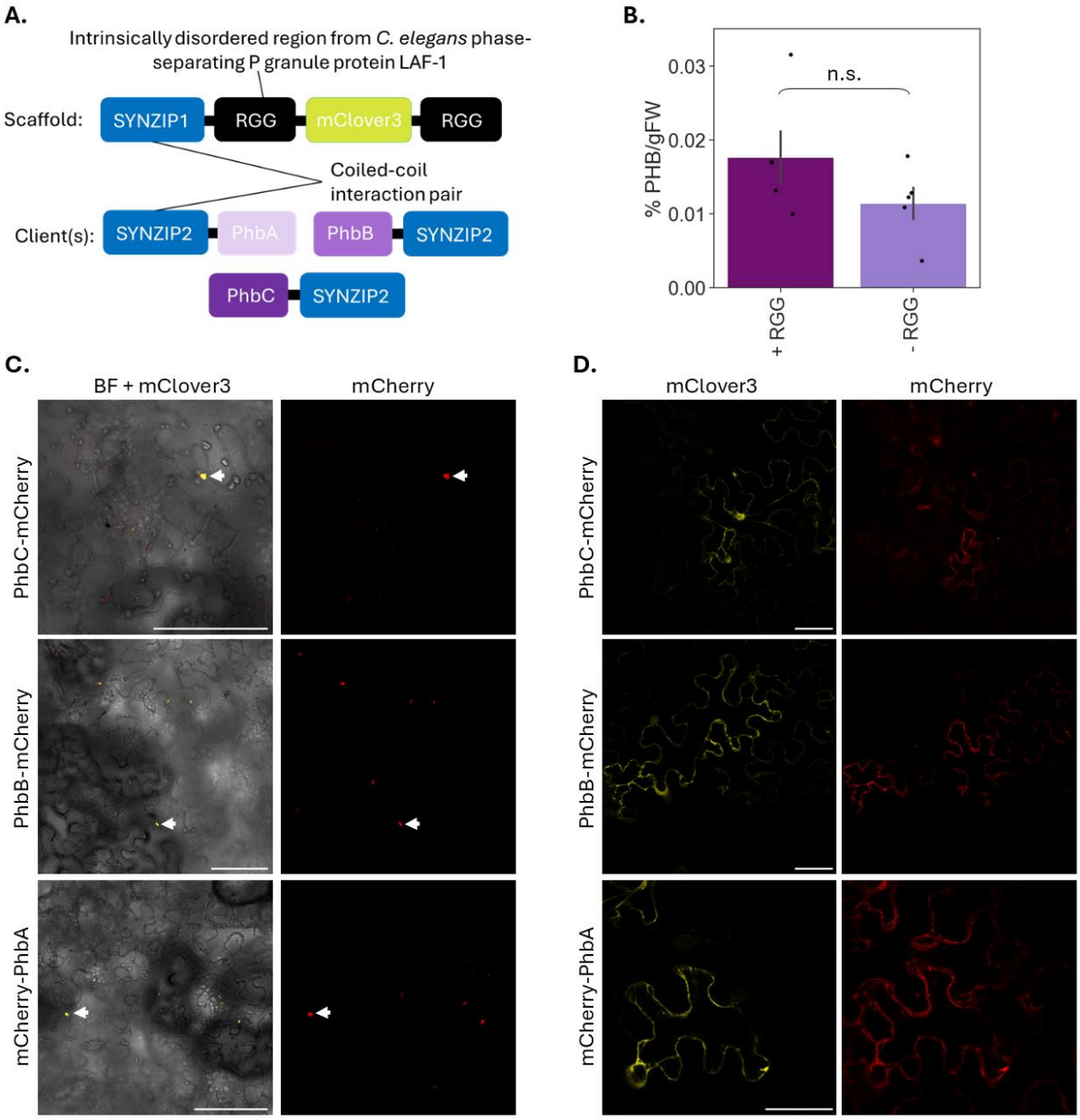


Figure 6.3: The localisation and activity of the three PHB-producing enzymes in synthetic biomolecular condensates in plants. (A) An illustration of the scaffold-client system designed and used to generate synthetic biomolecular condensates to which the three PHB-producing enzymes get recruited. PhbA, β-ketothiolase; PhbB, acetoacetyl-CoA reductase; PhbC, PHB synthase. (B) PHB contents in leaves 13 dpi with either the RGG- containing (+ RGG) or no-RGG-containing (- RGG) scaffold construct. Error bars represent SEM. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U). n = 5. (C) Co-localisation of mClover3 and mCherry-fused β-ketothiolase (mCherry-PhbA), acetoacetyl-CoA reductase (PhbB-mCherry), or PHB synthase (PhbC-mCherry). White arrow indicates a condensate. Scale bar is 100 µm. (D) The removal of the RGG domain ablates biomolecular condensate formation and results in diffuse cytosolic signal. Scale bar is 100 µm.

6.4.3. Combination of synthetic biomolecular condensate-formation with other metabolic engineering tools

I wanted to establish whether the RGG-based biomolecular condensates could be combined with other metabolic engineering tools, and whether the speed and ease of combinatorial construct testing enabled by *Agrobacterium* co-infiltration could be leveraged to give further insight into the production of cytosolic PHB in plants. To do this I cloned both the ACLY3KQ and *At*ALB versions of the ancillary constructs as described in section 5.4.1. into the pK7WG2 Gateway expression vector. I then carried out two co-infiltration experiments. In the first, I co-infiltrated the scaffold-client system for the recruitment of cytosolic *Md*CMS as described in section 6.4.1 with either blank agroinfiltration media or one of the two ancillary construct versions. In the second co-infiltration experiment, the scaffold-client system for the recruitment of all three PHB-pathway enzymes as described in section 6.4.2 was combined with the ancillary constructs in the same way. Leaf samples were collected and processed 5 dpi for citramalate quantification, and for PHB quantification leaves were re-infiltrated 5 dpi as described, and samples taken and processed 13 dpi.

The ancillary targets were chosen based on their role in the optimal route to cytosolic PHB accumulation as determined by the analysis of a model of the central metabolism of *M. polymorpha* (see Chapter 3). They were expected to increase cytosolic acetyl-CoA availability by encouraging the export of citrate from the vacuole and its conversion to acetyl-CoA by the overexpression of either the modified human ACLY3KQ or the chimeric *At*ALB (see section 5.4). As such I hypothesized that if the combination of these two targets was successful at increasing cytosolic acetyl-CoA

availability, the production of citramalate and PHB might also increase. There was a 1.12- to 1.69-fold increase in citramalate contents in leaves agroinfiltrated with the scaffold-client system as well as both version of the ancillary construct when compared to the scaffold-client system on its own (all $P > 0.05$); this was true of both the RGG- and non-RGG-containing scaffold versions (Figure 6.4A). There was also a higher (1.65-fold, $P = 0.0079$) citramalate content in leaves infiltrated with the RGG-containing version than the non-RGG-containing version of the scaffold-client system on their own.

When co-infiltrated with the scaffold-client system designed for PHB enzyme recruitment to the condensates, the ACLY3KQ version of the ancillary construct resulted in higher levels of PHB both with and without the RGG domains (3.12-fold and 2.47-fold, respectively, and $P = 0.032$ in both cases) (Figure 6.4B). Co-infiltration with the *At*ALB version resulted in an increase (2.25-fold, $P = 0.22$) with the RGG-containing scaffold but no increase with the scaffold not containing RGG. Similarly to the results shown in Figure 6.3B, there was a small increase (1.54-fold, $P = 0.55$) in PHB production when the scaffold-client system able to form biomolecular condensates was infiltrated compared to the no-RGG version. This was also true of co-infiltration with both ancillary construct versions between the RGG or non-RGG scaffold versions.

Taken together, this data indicates that the ancillary interventions are potentially resulting in higher acetyl-CoA availability in the cytosol. That they seem to have a greater impact on PHB production than on citramalate production might reflect the requirement of *Md*CMS for pyruvate as well as acetyl-CoA. In addition, localising enzymes to RGG-based synthetic biomolecular condensates seems to have some

benefit to product accumulation both in a single enzyme and a three-enzyme context.

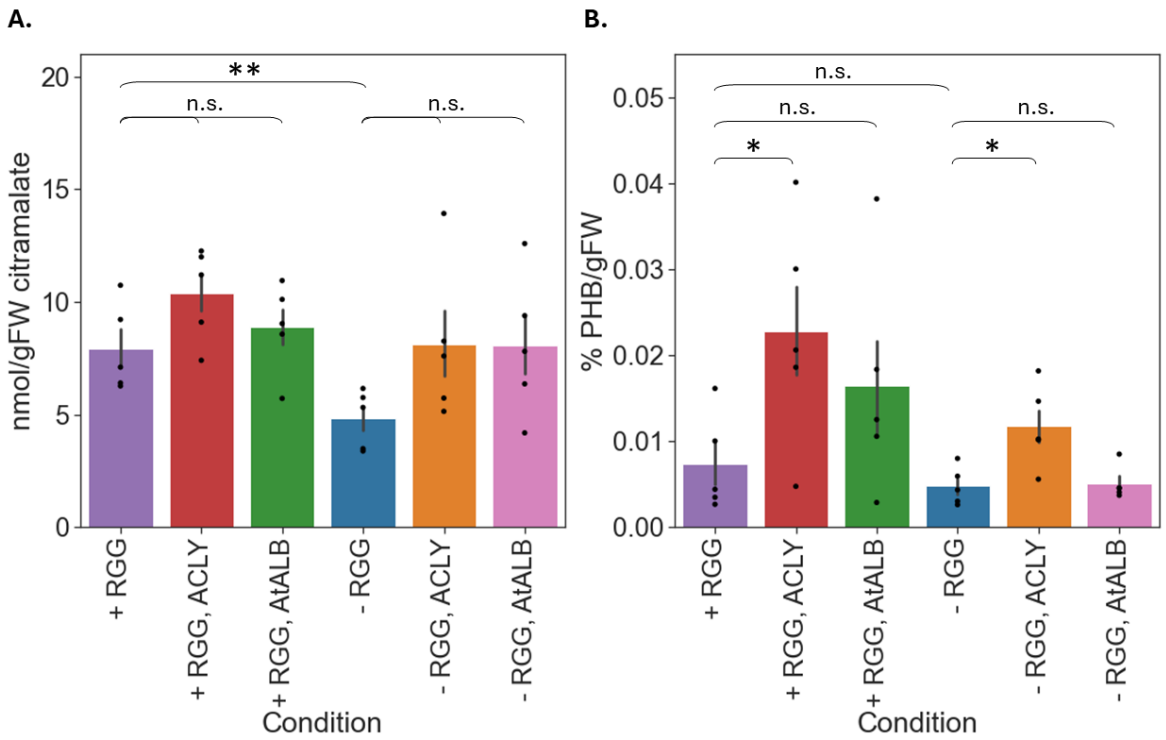


Figure 6.4: Insights into citramalate and PHB production in biomolecular condensates by the co-infiltration of ancillary metabolic interventions. (A) Citramalate contents in leaves 5 dpi with either the RGG- containing (+ RGG) or no-RGG-containing (- RGG) scaffold construct, and either the ACLY3KQ or *At*ALB version of the ancillary construct. Error bars represent SEM. n = 5. (B) PHB contents in leaves 13 dpi with either the RGG- containing (+ RGG) or no-RGG-containing (- RGG) scaffold construct, and either the ACLY3KQ or *At*ALB version of the ancillary construct. Error bars represent SEM. n = 5. In both panels '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05 (Mann-Whitney U).

6.5. Discussion

6.5.1. Insights into the biophysical properties of synthetic biomolecular condensates in plants

In this chapter I have described using an RGG-based scaffold-client system to build biomolecular condensates *in planta*. Although the ability of RGG domains to undergo LLPS *in vitro* has been well-established, not enough work has been done here to say whether the condensates constructed *in planta* are forming through LLPS and thereby whether they can be classed as phase-separated compartments. A published set of minimum experiments to support identification of LLPS *in vivo* suggested the following requirements: an observation of fusion events of spherical structures; demonstration of a concentration dependence of condensation (i.e. only assembly above c_{sat}); a test of whether LLPS-deficient mutants identified *in vitro* assemble in cells (*280*). Clearly the RGG domain is essential for the formation of biomolecular condensates *in planta* given that biomolecular condensate formation is completely ablated in its absence (Figures 6.1B, 6.2D, 6.3D), and this supports *in vitro* results showing that the RGG domain is necessary and sufficient for phase-separation (*103*). However, neither the fusion of spherical droplets nor concentration-dependent condensate formation were characterised in this work.

The identification of spherical droplets that can fuse is used as a hallmark of LLPS because liquids are known to form droplets, fuse, coalesce and drip as a result of their surface tension (*280*). In the work undertaken here some biomolecular condensates appeared spherical (Figure 6.1C) but no comprehensive quantification of circularity

was undertaken, and the fusion of droplets was not observed. In addition, clearly not all biomolecular condensates were spherical and large irregularly shaped condensates were observed in the cytosol (Figure S6.1C). Although the formation of a condensed phase through LLPS is usually reversible, condensates can also undergo further transitions such as to gels or hydrogels. The formation of these states is often irreversible, and these condensates are resistant to dissolution, often requiring high salt or denaturant. The large and irregularly-shaped cytosolic biomolecular condensates mentioned were observed from 7 dpi, and could represent more solid-like condensates especially given that changes in the physical properties of biomolecular condensates such as liquid-gel transitions are known to become more important with time and condensate maturation (*99*). FRAP analysis of these structures might give some insight into the mobility of these structures and thereby their physical state, given that the diffusion of components is arrested in gel-like states (*280*). Alternatively, super-resolution imaging could be used to determine whether the concentration of fluorescent molecules is uniformly and randomly distributed in these structures, as would be expected in a liquid state (*91*).

Recovery from photobleaching as quantified using FRAP is often used as a criterion for defining a 'liquid-like' phase-separated structure. Recent work, however, has highlighted that recovery rates from FRAP have many different possible interpretations. For instance, some non-liquid structures, such as spongy structures which transiently bind the species whose recovery is being monitored, can have fast FRAP recovery rates. Even proteins shown to engage in high-affinity and stable binding interactions can show recovery in FRAP (*288*). Conversely, structures otherwise

unequivocally demonstrated as forming as a result of LLPS can have relatively slow recovery rates (*92*). This is because the recovery of fluorescence is dependent on complex interactions between many parameters including the diffusion coefficient and concentration of the molecule being measured, the association with any binding partners, but also on technical considerations regarding the microscope and detectors being used (*289*) or size of the bleached area (*290*). This has resulted in recovery times for droplets generated from the same IDR ranging from sub-seconds to minutes (*91*). In addition, results from half- or full-FRAP are often used interchangeably, even though they quantify very different dynamics. A full-FRAP provides information regarding the influx of fluorescent molecules from the outside of the biomolecular condensate but only a half-FRAP can provide information regarding internal rearrangements of the biomolecular condensate and thereby its biophysical state (*91*).

In this work, full-FRAP revealed that fluorescence recovered to approximately 45 % of its pre-FRAP level, the recovery time was 10.3-27.3 s, and the calculated diffusion constant was similar to published values (see section 6.3 and Figure 6.1E). Using this information, and given that two out of the three diagnostic experiments as mentioned have not been carried out, I cannot make a definitive statement as to whether the condensate is a liquid. Instead, I can say that the biomolecular condensates are composed of a mobile and an immobile fraction, and that the mobile fraction is being exchanged with the outside of the biomolecular condensate relatively quickly.

6.5.2. The amenability of synthetic biomolecular condensates to metabolic engineering

The formation of a biomolecular condensate necessarily entails a diffusion barrier between the inside and the outside of the condensate. This results in an energetic barrier to entering or leaving the biomolecular condensate (*91*). However, no comprehensive framework exists with which to predict whether or how a given metabolite will diffuse into and out of a given condensate. Given the detection of both citramalate and PHB when *Md*CMS or the PHB-producing enzymes were targeted to condensates *in planta* we can conclude that acetyl-CoA (and probably pyruvate, H_2O, and NADPH given that they are also involved in the reactions) is able to effectively diffuse into the condensate. Single particle tracking has previously been used to visualise the diffusion dynamics into and inside biomolecular condensates (*291*), and could be used here to confirm the diffusion of various molecules into the biomolecular condensate. It is of course possible that at least some of the end-product is the result of activity of enzymes expressed but not recruited to the biomolecular condensate. Given that levels of citramalate and PHB were consistently either the same or higher in leaves infiltrated with the construct containing the RGG domains (and thereby able to form biomolecular condensates to which the enzymes could be recruited) (Figures 6.2B, 6.3B, 6.4A-B), it is likely that at least some of the detected citramalate and PHB were produced from biomolecular condensate-recruited enzymes. Nonetheless, further work should be undertaken to quantify the relative amounts of enzyme recruited to the biomolecular condensate versus cytosolic enzyme.

Experiments involving the immobilisation of NanoLuc inside an RGG-based

condensate *in vitro* was able to show that the reaction between NanoLuc and its substrate coelenterazine was initially predominantly occurring at the water-condensate interface, and gradually reached the inside of the biomolecular condensate as the substrate diffused into the condensate (*104*). This would imply that smaller condensates may be more advantageous for product accumulation given their higher surface area to volume ratio. It also shows that enzymatic activity and product production may be heterogeneously distributed throughout the condensate. Given that PHB accumulates in insoluble granules, it would be interesting to determine where these granules form. If most enzymatic activity occurs on the condensate boundary, then the accumulation of PHB granules might be concentrated there. It is also unclear how the insoluble granule would partition within the protein-based condensate, given that the polymer is highly hydrophobic but the RGG IDRs are charged.

Compartmentation of metabolism through biomolecular condensates can impact metabolism in two main ways. Firstly, recent work has demonstrated in detail that kinetic benefits can be imparted on enzymes localized to such condensates (*78*), either by mass action kinetics or by encouraging probabilistic metabolite channelling (see section 1.5). As mentioned, the production levels of citramalate and PHB in leaves where enzymes were recruited to biomolecular condensates were similar or higher than levels obtained when enzymes were purely cytosolic. It may be that the higher levels of product accumulation are due to kinetic benefits, but further experiments would be needed to prove this. For instance, in the case of PHB production, SYNZIP2 could be removed from acetoacetyl-CoA reductase thereby ablating its recruitment to

the condensate and effects on PHB production quantified. Secondly, the compartmentation of specific enzymes together can provide control over metabolic flux, for example at metabolic branch points. In this work a common theme has been the competition between endogenous metabolism and engineered PHB production for acetyl-CoA. In this chapter and in section 5.4.2 I have demonstrated that the co-expression of a construct encouraging vacuolar citrate export and conversion to acetyl-CoA can encourage PHB production. The next step would be to localise ACLY3KQ or *At*ALB to the biomolecular condensate thereby providing the PHB pathway with its own co-localised source of acetyl-CoA.

In this work I was able to build synthetic biomolecular condensates in transiently transformed *N. benthamiana* leaves. To my knowledge, this is the first time such structures have been built in plant cells. However, no biomolecular condensates were observed when the same construct was transformed into *M. polymorpha*. Certain genetic sequences are known to be interpreted very differently depending on the organism in which they are introduced. For example, the triple HA epitope acts as a vacuolar sorting signal in *M. polymorpha* thallus cells but is widely used in other plants for tagging non-vacuolar proteins (*292*). It is also known that some genes are difficult to express in *M. polymorpha* (*293*). Addition of the 5'-UTR of the liverwort homolog of tobacco ADH-like UDP glucose dehydrogenase (used to increase gene expression in tobacco and *Arabidopsis*) is effective at increasing transgene expression in *M. polymorpha* and could be used here (*294*). It is also possible that the constitutive expression of the scaffold-client system is detrimental to plant health, and so the transgenes are either silenced and/or only plants with insignificant expression levels

survive. If this is the case, this presents a considerable obstacle to using this system as a tool in plant metabolic engineering.

7. Discussion

7.1. Summary of main findings

The establishment of a plant-based production system with high product titre is exceedingly difficult. This is at least partly due to the complexity, unpredictability, and tight regulation of central metabolism such that metabolic engineering interventions often have unexpected and undesirable effects. As a result, establishing a plant system with economically viable levels of product accumulation often requires multiple rounds of engineering. This iterative improvement is common in the engineering of microorganisms, but in plants is limited by the slow and laborious process of generating transgenic plants. *M. polymorpha* has a haploid-dominant lifecycle meaning that homozygous lines can be generated relatively quickly and easily. Combined with its ease of propagation and recently established transformation techniques, this has resulted in its emergence as a possible modern testbed species for plant metabolic engineering.

The aim of this research project was to establish whether *M. polymorpha* is indeed a useful testbed species for metabolic engineering, using the production of PHB as a case study. The production of PHB in plants presents challenges in terms of low substrate availability, low product titre, and complex interactions with endogenous metabolism. For example, it is widely hypothesized that the production of PHB competes with biomass-producing reactions for acetyl-CoA, leading to stunted plants. These issues are generally applicable to plant metabolic engineering, making

PHB production a good target pathway for the development of plant metabolic engineering techniques.

Metabolic modelling can be used to determine and rapidly test the effects of metabolic engineering strategies on the entire metabolic network, and thereby help understand possible complex and unexpected consequences of an engineering intervention. I constructed a computational model of *M. polymorpha* metabolism and used it to inform the metabolic engineering strategy for PHB production. I established that the cytosol is, in principle, a more promising location for PHB engineering, and identified specific engineering targets which might increase the capacity of the existing metabolic network to support maximal PHB production. These included the vacuolar citrate-malate exchanger *At*tDT, and the enzyme ACL in the form of either *At*ALB or ACLY3KQ. To characterise the effect of these targets on PHB accumulation, I generated *M. polymorpha* plants capable of constitutive PHB production. The co-expression of the computationally identified targets in this background did result in enhanced yield of PHB, but trade-offs with biomass production still existed as super-transformed lines were significantly smaller.

To improve the yield of PHB by reducing the impact on endogenous plant metabolism, I implemented inducible PHB biosynthesis in *M. polymorpha* using the PULSE optogenetic system. Inducible expression is often used in metabolic engineering to produce molecules known to negatively affect the host, due to its ability to temporally separate product accumulation and 'normal' growth. Despite resulting in a developmental phenotype, the optogenetic system was fully functional in *Marchantia* with levels of induction and leakiness comparable to published work in *Arabidopsis*.

The PULSE system could also inducibly control the expression of the three PHB pathway enzymes, resulting in PHB levels comparable to constitutively producing plants. The inducible production of PHB had different effects on plant growth depending on the developmental stage of the tissue in which it was induced. In gemmae, the inducible production of PHB did not result in a growth benefit, whereas in thallus cuttings there was a clear advantage to inducible production of PHB over constitutive production in terms of the negative effect on growth. The beneficial effects of inducible PHB production might have been tempered by the induction conditions used, given that monochromatic red light was shown to encourage plant growth.

Finally, I investigated the targeting of PHB pathway enzymes to biomolecular condensates for enhanced PHB production. To achieve this I designed, built, and characterised synthetic biomolecular condensates in transiently transformed *N. benthamiana* plants. Although these structures were built from the RGG IDR known to phase-separate *in vitro*, not enough evidence was presented here to conclude that the biomolecular condensates formed through LLPS. Using a coiled-coil dimerising pair I was able to recruit both mCherry and one (*Md*CMS) or several (PHB pathway) enzymes to the condensates and showed that the recruited enzymes retained their functionality. This implied that the condensates were permeable to the precursors for the reactions they catalysed, namely acetyl-CoA, pyruvate, NADPH and H_2O. However, I was unable to generate synthetic biomolecular condensates in stably transformed *Marchantia* lines. This problem needs to be overcome if synthetic biomolecular condensates are to become a useful tool in plant metabolic engineering.

7.2. The production of PHB in *M. polymorpha*

In this work, PHB was produced constitutively, inducibly, in combination with computationally derived targets, and in synthetic biomolecular condensates. The highest yield was obtained in constitutive PHB-producing plants, where an average yield of 0.074 %PHB gFW^{-1} was achieved for line #3 grown for 39 d (Figure 5.1E). Comparisons against published values for cytosolic PHB yields in plants are complicated by the reporting of results using different units (e.g. per gram dry weight [gDW^{-1}] or gFW^{-1}) or the reporting of yields in specific tissues or organs (e.g. leaves). This can result in inflated yield values if heterogeneous production is expected for example when producing PHB in plastids (given that plastid contents in different organs can vary considerably (*295*)), or when using promoters which show organ-dependent expression levels (*296*). In this work, PHB values were reported as %PHB gFW^{-1} of the whole plant. Even when whole plants were not collected (because they were too big), samples were taken such that they represented the different stages of thallus growth and PHB contents in the sample was representative of PHB contents in the whole plant. The accepted estimate is that PHB must accumulate to 10-12.5% per unit whole-plant dry weight for the process to be cost-competitive with petroleum-derived plastics (*46, 47*). To convert the %PHB gFW^{-1} achieved using constitutive expression in this work to %PHB gDW^{-1}, a dry weight to fresh weight ratio for *M. polymorpha* of 0.0547 was estimated from data in (*297*), and this resulted in a maximal PHB yield of 1.43 %PHB gDW^{-1} of the whole plant. Although much lower than the minimum cost-competitive yield, this is higher than the highest yield of cytosolic PHB reported to date in a terrestrial plant, which is 0.8 % per unit leaf dry weight (converted

from value reported as % per unit leaf fresh weight using conversion factor from (*202*))(*65*).

There therefore seems to exist an inherent benefit to cytosolic PHB production in *M. polymorpha* over other terrestrial land plant species in which it has been attempted. Bryophytes produce high levels of very long chain polyunsaturated fatty acids (LCPFAs) such as arachidonic acid and eicosapentaenoic acid which are uncommon in higher plants (*298*). In fact, a homology search of *Arabidopsis* genes in the *Marchantia* genome used for the generation of the model of *M. polymorpha* central metabolism resulted in the addition of 57 reactions and 56 metabolites to account for long-chain fatty acids present in *Marchantia* but not in *Arabidopsis* (section 2.1.1). The biosynthesis of LCPFAs in bryophytes begins in the plastid with the help of plastidial acetyl-CoA and malonyl-CoA, and 16:0, 18:0 and 18:1 chains are then transferred to the endoplasmic reticulum for production of LCPFAs by several fatty acid desaturases (*298*). It is possible that a metabolic network able to support large amounts of LCPFA biosynthesis might also more easily accommodate the biosynthesis of PHB. Bryophytes are also known to produce large amounts of terpenoids, of which more than 1600 have been identified in liverworts (*299*). In plants, terpenoids can be synthesized via one of two compartmentalised pathways. Some terpenoids are synthesized in the plastid by the condensation of pyruvate and glyceraldehyde-3-phosphate. However, the mevalonic acid pathway for terpenoid biosynthesis is mostly cytosolic and involves the condensation of three acetyl-CoA units leading to the synthesis of 3-hydroxy-3-methylglutaryl-CoA (*300*). Because of their propensity for terpenoid accumulation it is possible that bryophytes have a higher cytosolic acetyl-

CoA concentration than other land plants and/or possess a heightened ability for acetyl-CoA biosynthesis.

Due to its small size and slow growth, *M. polymorpha* has never been touted as a potential commercial production host and *P. patens* is the only bryophyte used for industrial biotechnological purposes (*298*). *P. patens* has been suggested as a production host for sesquiterpenoids and diterpenoids, and is already used in the perfume industry for the production of patchoulol (*44*). The work presented here shows that *P. patens* could be considered a production host for other acetyl-CoA-derived products too, such as PHB.

7.3. Insights into metabolic bottlenecks to plant PHB production

Published work regarding the production of PHB in plants report clear negative effects on phenotype with higher PHB accumulation levels (see Table 1.1). In many cases this involves negative effects on growth, and this was also shown in this work where strong negative correlations existed between final plant size and PHB accumulation levels in plants constitutively producing PHB (Figures 5.1B and 5.4B). It has been widely speculated that part of the reason for this is the competition between PHB- and biomass-producing reactions (such as fatty acid synthesis) for acetyl-CoA, but to my knowledge this has never been empirically determined. It has been shown, however, that phenotypes of PHB-producing *Arabidopsis* plants resemble some phenotypes present in plants with downregulated cytosolic acetyl-CoA synthesis (*301*). It has also been shown that the overexpression of ACL in an *Arabidopsis* background producing

PHB can rescue negative growth phenotypes associated with PHB production (*61*). In this work, computational modelling of cytosolic PHB production in *M. polymorpha* showed that high flux through ACL was essential for optimal PHB production. In addition, I showed that co-expressing two computationally derived targets expected to increase cytosolic acetyl-CoA availability resulted in higher yields of PHB when expressed as a %PHB gFW^{-1}, both in stable transgenic lines and in transiently transformed *N. benthamiana* leaves. Taken together these data suggest that acetyl-CoA does have a role in limiting simultaneous biomass and PHB production in plants.

There are reasons other than competition for acetyl-CoA, however, that could be affecting plant health. The most obvious of these is toxicity of the PHB granules themselves. The PHB polymer is expected to be inert, but there has been speculation that its very high hydrophobicity might cause problems for the cell and that this is part of the reason PHB granules are coated in phasin proteins in bacteria (*302*). However, when the phasin protein PhaP was co-expressed with the PHB-producing enzymes in *Arabidopsis* it was not able to rescue plant growth defects despite correctly localising to the surface of the PHB granules (*71*). Alternatively, the intermediates in PHB biosynthesis, namely acetoacetyl-CoA and/or hydroxybutyryl-CoA, could be accumulating to toxic levels, which could also be symptomatic of inefficient flux through the heterologous PHB pathway. The accumulation of reaction intermediates when expressing a heterologous pathway is relatively common, and can be the result of unbalanced enzyme stoichiometry (*303*), but to my knowledge no work has been undertaken to characterise this in stable PHB-producing transgenic plants. In *E. coli*, the sensitivity of PHB production to perturbations in PHB pathway gene expression

was analysed by the sequential overexpression of one of the three genes in a background expressing the entire PHB operon. It was found that acetoacetyl-CoA reductase exerted the most flux control over the system whilst the overexpression of either β-ketothiolase or PHB synthase had no effect on final yield (*304*). This finding supports previous *in vitro* and *in silico* work on *E. coli* showing enzyme activity to be the lowest (*305*) and the flux control coefficient to be the highest (*306*) for acetoacetyl-CoA reductase. The work that comes the closest to quantifying effects of enzyme stoichiometry on plant-based PHB production used tobacco BY-2 cells to show that an engineered version of acetoacetyl-CoA with a 3.6-fold higher k_{cat} resulted in approximately 3-fold higher PHB production (*307*), suggesting this enzyme to be rate-limiting.

Another reason for the possible negative effects on plants associated with PHB production is competition for NADPH, and various aspects of the work presented here support this conclusion. For instance, when analysing a computational model of cytosolic PHB production in *Marchantia*, I showed that there was a perfect positive correlation between flux to cytosolic PHB accumulation and the GAPN reaction. This reaction catalyses the oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate using the reduction of $NADP^+$ to NADPH. Presumably this reaction is favoured in order to provide sufficient NADPH for acetoacetyl-CoA reductase. In addition, a metabolomic analysis of the differences in metabolite profile between WT Tak2 plants and constitutive PHB-producing plants showed there to be a significant decrease in tyrosine in the PHB-producing plants. Given the requirement for NADPH in the biosynthesis of tyrosine, this would provide further evidence of the importance of

NADPH in ensuring simultaneous PHB production and normal endogenous metabolism. It has previously been suggested, by the analysis of a stoichiometric model of *S. cerevisiae* expressing the three enzymes required for PHB synthesis during slow growth, that the theoretical yield of PHB could be improved by including a combination of ACL and a transhydrogenase reaction to catalyse the conversion of NADH and $NADP^+$ to NAD^+ and NADPH, respectively (*194*). Taken together this evidence suggests that increasing cytosolic NADPH provision should be considered as an additional engineering target in the production of cytosolic PHB.

7.4. Is *M. polymorpha* a good testbed species for metabolic engineering?

In this project I wanted to determine whether *M. polymorpha* could be a useful testbed species for metabolic engineering. Using the production of PHB as a case study, I have shown that optogenetic induction can be used in *M. polymorpha* and that in some cases it can result in higher yields of PHB. I have also shown that a computational model can be built for *M. polymorpha* and can be used to identify targets which, when implemented in a plant background constitutively producing PHB, result in higher yields. Both of these approaches have been effective as a way of exploring how to improve plant based PHB yield and have given insights into the interplay between PHB accumulation and endogenous metabolism. Using *Marchantia* has enabled me to rapidly generate homozygous transgenic lines and has allowed for easy propagation using gemmae or thallus cuttings even when developmental phenotypes were observed.

However, although I successfully built synthetic biomolecular condensates in transiently transformed *N. benthamiana* leaves, I was unable to do this is *Marchantia*. Whether this is an experimental issue that can be overcome or whether such structures cannot be built in *Marchantia* remains to be seen. Other challenges include silencing observed in some transgenic lines as a result of thallus fragment subculture. Although presenting an obstacle to the maintenance of stable *Marchantia* lines, being able to propagate lines using vegetative tissue means that mutations leading to sterility can be more readily investigated as there is no reliance on seed production. In addition, although the small genome size of *Marchantia* makes it easier to engineer, the resulting lack of genetic and functional redundancy also results in a bigger potential for interactions between engineered and endogenous metabolism as was hypothesized to be occurring between the *Arabidopsis* PhyB-PIF system forming part of PULSE and the *Mp*Phy-*Mp*PIF pair (Chapter 4).

Finally, the concept of a testbed species relies on the idea that tools and techniques developed for such a species can be directly translated to a production system. In this way it is an extension of the concept of a model species, which refers to a species that is usually small, easy to grow, has a short generation time, and is amenable to genetic manipulation, and which is studied with the expectation that it will provide insights into other species (*308*). The same attributes which make a plant species a good model or testbed, however, preclude its choice as a production plant. For instance, one of the most important features of a production crop is fast and high levels of biomass production (*309*) whereas testbed species are usually small. As such the commercial application of a strategy developed in such a species necessarily involves the

transition to a production plant. However, even though central metabolism is conserved in the green lineage (*140*), it is likely that there will be substantial differences in the regulation of, and demands on, that central metabolism in different species. This means that engineering approaches developed in a testbed species might still have to be significantly modified to work in a production plant. Arguably, then, efforts should be directed at finding a plant species which possesses the advantages of working with *Marchantia* (fast and easy to transform), but which also grows fast and to a large size, and thereby doing away with the need for testbed species at all.

References

1. J. Choi, S. Y. Lee, Factors affecting the economics of polyhydroxyalkanoate production by bacterial fermentation. *Applied Microbiology and Biotechnology* **51**, 13-21 (1999).
2. D. Tuse, T. Tu, K. A. McDonald, Manufacturing economics of plant-made biologics: case studies in therapeutic and industrial enzymes. *Biomed Research International*, (2014).
3. M. L. Yang *et al.*, Accumulation of high-value bioproducts *in planta* can improve the economics of advanced biofuels. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 8639-8648 (2020).
4. J. A. N. Brophy *et al.*, Synthetic genetic circuits as a means of reprogramming plant roots. *Science* **377**, 747-+ (2022).
5. L. J. Sweetlove, J. Nielsen, A. R. Fernie, Engineering central metabolism - a grand challenge for plant biologists. *The Plant Journal* **90**, 749-763 (2017).
6. E. M. Zhao *et al.*, Light-based control of metabolic flux through assembly of synthetic organelles. *Nature Chemical Biology* **15**, 589-+ (2019).
7. E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A. L. Barabási, Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature* **427**, 839-843 (2004).
8. J. Nielsen, J. D. Keasling, Engineering cellular metabolism. *Cell* **164**, 1185-1197 (2016).
9. L. J. Sweetlove, D. Fell, A. R. Fernie, Getting to grips with the plant metabolic network. *Biochemical Journal* **409**, 27-41 (2007).
10. L. Bender-Machado *et al.*, Expression of a yeast acetyl CoA hydrolase in the mitochondrion. *Plant Molecular Biology* **55**, 645-662 (2004).
11. S. Wu *et al.*, Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology* **24**, 1441-1447 (2006).
12. S. Di Fiore *et al.*, Targeting tryptophan decarboxylase to selected subcellular compartments of tobacco plants affects enzyme stability and *in vivo* function and leads to a lesion-mimic phenotype. *Plant Physiology* **129**, 1160-1169 (2002).
13. I. F. Kappers *et al.*, Genetic engineering of terpenoid metabolism attracts bodyguards to *Arabidopsis*. *Science* **309**, 2070-2072 (2005).

14. J. H. Lynch, X.-Q. Huang, N. Dudareva, Silent constraints: the hidden challenges faced in plant metabolic engineering. *Current Opinion in Biotechnology* **69**, 112-117 (2021).

15. U. Heinig, M. Gutensohn, N. Dudareva, A. Aharoni, The challenges of cellular compartmentalization in plant metabolic engineering. *Current Opinion in Biotechnology* **24**, 239-246 (2013).

16. M. Farhi *et al.*, Generation of the potent anti-malarial drug artemisinin in tobacco. *Nature Biotechnology* **29**, 1072-1074 (2011).

17. S. Kumar *et al.*, Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metabolic Engineering* **14**, 19-28 (2012).

18. B. Saxena *et al.*, Metabolic engineering of chloroplasts for artemisinic acid biosynthesis and impact on plant growth. *Journal of Biosciences* **39**, 33-41 (2014).

19. N. Vemuri Goutham, A. Aristidou Aristos, Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. *Microbiology and Molecular Biology Reviews* **69**, 197-216 (2005).

20. M. S. Kwon, B. T. Lee, S. Y. Lee, H. U. Kim, Modeling regulatory networks using machine learning for systems metabolic engineering. *Current Opinion in Biotechnology* **65**, 163-170 (2020).

21. S. Yilmaz, A. Nyerges, J. van der Oost, G. M. Church, N. J. Claassens, Towards next-generation cell factories by rational genome-scale engineering. *Nature Catalysis* **5**, 751-765 (2022).

22. K. V. Presnell, H. S. Alper, Systems metabolic engineering meets machine learning: a new era for data-driven metabolic engineering. *Biotechnology Journal* **14**, 1800416 (2019).

23. C. S. Pikaard, O. Mittelsten Scheid, Epigenetic regulation in plants. *Cold Spring Harbor Perspectives in Biology* **6**, (2014).

24. C.-Y. Huang, H. Jin, Coordinated epigenetic regulation in plants: a potent managerial tool to conquer biotic stress. *Frontiers in Plant Science* **12**, (2022).

25. A. Movahedi, J. Zhang, R. Amirian, Q. Zhuge, An efficient *Agrobacterium*-mediated transformation system for poplar. *International Journal of Molecular Sciences*. 2014 (10.3390/ijms150610780).

26. F. Sainsbury, E. C. Thuenemann, G. P. Lomonossoff, pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* **7**, 682-693 (2009).

27. E. D. Carlson, J. Rajniak, E. S. Sattely, Multiplicity of the *Agrobacterium* infection of *Nicotiana benthamiana* for transient DNA delivery. *ACS Synthetic Biology* **12**, 2329-2338 (2023).

28. B. J. Schultz, S. Y. Kim, W. Lau, E. S. Sattely, Total biosynthesis for milligram-scale production of etoposide intermediates in a plant chassis. *Journal of the American Chemical Society* **141**, 19231-19235 (2019).

29. A. Fulton, H. Lai, Q. Chen, C. Zhang, Purification of monoclonal antibody against *Ebola* GP1 protein expressed in *Nicotiana benthamiana*. *Journal of Chromatography A* **1389**, 128-132 (2015).

30. Y. Zhang *et al.*, A highly efficient *Agrobacterium*-mediated method for transient gene expression and functional studies in multiple plant species. *Plant Communications* **1**, 100028 (2020).

31. K. Norkunas, R. Harding, J. Dale, B. Dugdale, Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods* **14**, 71 (2018).

32. F. Grosse-Holz *et al.*, The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire. *Plant Biotechnology Journal* **16**, 1068-1084 (2018).

33. A. Horn *et al.*, Natural products from bryophytes: from basic biology to biotechnological applications. *Critical Reviews in Plant Sciences* **40**, 191-217 (2021).

34. A. Kubota, K. Ishizaki, M. Hosaka, T. Kohchi, Efficient *Agrobacterium*-mediated transformation of the liverwort *Marchantia polymorpha* using regenerating thalli. *Bioscience Biotechnology and Biochemistry* **77**, 167-172 (2013).

35. K. Ishizaki, S. Chiyoda, K. T. Yamato, T. Kohchi, *Agrobacterium*-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. *Plant and Cell Physiology* **49**, 1084-1091 (2008).

36. D. Schaefer, J. P. Zryd, C. D. Knight, D. J. Cove, Stable transformation of the moss *Physcomitrella patens*. *Molecular and General Genetics MGG* **226**, 418-424 (1991).

37. C. Collonnier *et al.*, CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss *Physcomitrella patens*. *Plant Biotechnology Journal* **15**, 122-131 (2017).

38. S. S. Sugano *et al.*, CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. *Plant and Cell Physiology* **55**, 475-481 (2014).

39. A. Peramuna *et al.*, Evaluation of synthetic promoters in *Physcomitrella patens*. *Biochemical and Biophysical Research Communications* **500**, 418-422 (2018).

40. S. Sauret-Gueto *et al.*, Systematic tools for reprogramming plant gene expression in a simple model, *Marchantia polymorpha*. *ACS Synthetic Biology* **9**, 864-882 (2020).

41. N. J. Patron *et al.*, Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist* **208**, 13-19 (2015).

42. K. Ishizaki *et al.*, Development of Gateway binary vector series with four different selection markers for the liverwort *Marchantia polymorpha*. *PLOS One* **10**, (2015).

43. N. K. B. Khairul Ikram *et al.*, Stable production of the antimalarial drug artemisinin in the moss *Physcomitrella patens*. *Frontiers in Bioengineering and Biotechnology* **5**, (2017).

44. X. Zhan, Y.-H. Zhang, D.-F. Chen, H. T. Simonsen, Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and α/β-santalene. *Frontiers in Plant Science* **5**, (2014).

45. M. Takemura, H. Kanamoto, S. Nagaya, K. Ohyama, Bioproduction of prostaglandins in a transgenic liverwort, *Marchantia polymorpha*. *Transgenic Research* **22**, 905-911 (2013).

46. H. Lu *et al.*, Reconfiguring plant metabolism for biodegradable plastic production. *BioDesign Research* **2020**, 9078303 (2020).

47. M. N. Somleva, O. P. Peoples, K. D. Snell, PHA bioplastics, biochemicals, and energy from crops. *Plant Biotechnology Journal* **11**, 233-252 (2013).

48. J. C. Mortimer, Plant synthetic biology could drive a revolution in biofuels and medicine. *Experimental Biology and Medicine* **244**, 323-331 (2019).

49. C. G. de Oliveira Dal'Molin, L.-E. Quek, P. A. Saa, R. Palfreyman, L. Keld Nielsen, From reconstruction to C-4 metabolic engineering: A case study for overproduction of polyhydroxybutyrate in bioenergy grasses. *Plant Science* **273**, 50-60 (2018).

50. E. Okamura, T. Tomita, R. Sawa, M. Nishiyama, T. Kuzuyama, Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11265-11270 (2010).

51. Y. Poirier, D. E. Dennis, K. Klomparens, C. Somerville, Polyhydroxybutyrate, a biodegradable thermoplastic, produced in transgenic plants. *Science* **256**, 520-523 (1992).

52. K. Bohmert *et al.*, Transgenic *Arabidopsis* plants can accumulate polyhydroxybutyrate to up to 4% of their fresh weight. *Planta* **211**, 841-845 (2000).

53. E. Agathokleous, M. C. Rillig, J. Peñuelas, Z. Yu, One hundred important questions facing plant science derived using a large language model. *Trends in Plant Science* **29**, 210-218 (2024).

54. J. Dobrogojski, M. Spychalski, R. Luciński, S. Borek, Transgenic plants as a source of polyhydroxyalkanoates. *Acta Physiologiae Plantarum* **40**, 162 (2018).

55. X. Yang *et al.*, Plant biosystems design research roadmap 1.0. *BioDesign Research* **2020**, 8051764 (2020).

56. C. Nawrath, Y. Poirier, C. Somerville, Targeting of the polyhydroxybutyrate biosynthetic-pathway to the plastids of *Arabidopsis thaliana* results in high-levels of polymer accumulation. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12760-12764 (1994).

57. K. Bohmert-Tatarev, S. McAvoy, S. Daughtry, O. P. Peoples, K. D. Snell, High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. *Plant Physiology* **155**, 1690-1708 (2011).

58. A. Ali *et al.,* WIPO, Ed. (2012), chap. WO2012037324A3037324A3.

59. R. B. McQualter *et al.*, The use of an acetoacetyl-CoA synthase in place of a β-ketothiolase enhances poly-3-hydroxybutyrate production in sugarcane mesophyll cells. *Plant Biotechnology Journal* **13**, 700-707 (2015).

60. D. J. Oliver, B. J. Nikolau, E. S. Wurtele, Acetyl-CoA-Life at the metabolic nexus. *Plant Science* **176**, 597-601 (2009).

61. S. F. Xing *et al.*, ATP citrate lyase activity is post-translationally regulated by sink strength and impacts the wax, cutin and rubber biosynthetic pathways. *The Plant Journal* **79**, 270-284 (2014).

62. S. Bordel, R. J. M. van Spanning, F. Santos-Beneit, Imaging and modelling of poly(3-hydroxybutyrate) synthesis in *Paracoccus denitrificans*. *AMB Express* **11**, 113 (2021).

63. R. B. McQualter *et al.*, Systems biology and metabolic modelling unveils limitations to polyhydroxybutyrate accumulation in sugarcane leaves; lessons for C-4 engineering. *Plant Biotechnology Journal* **14**, 567-580 (2016).

64. Y. Suzuki *et al.*, Enzyme inhibitors to increase poly-3-hydroxybutyrate production by transgenic tobacco. *Bioscience Biotechnology and Biochemistry* **66**, 2537-2542 (2002).

65. Y. Poirier, C. Somerville, L. A. Schechtman, M. M. Satkowski, I. Noda, Synthesis of high-molecular-weight poly(R -(-)-3-hydroxybutyrate) in transgenic *Arabidopsis thaliana* plant-cells. *International Journal of Biological Macromolecules* **17**, 7-12 (1995).

66. L. A. Petrasovits, M. P. Purnell, L. K. Nielsen, S. M. Brumbley, Production of polyhydroxybutyrate in sugarcane. *Plant Biotechnology Journal* **5**, 162-172 (2007).

67. G. Corrado, M. Karali, Inducible gene expression systems and plant biotechnology. *Biotechnology Advances* **27**, 733-743 (2009).

68. E. C. F. Forestier, A. C. Cording, G. J. Loake, I. A. Graham, An engineered heat-inducible expression system for the production of casbene in *Nicotiana*

benthamiana. *International Journal of Molecular Sciences*. 2023 (10.3390/ijms241411425).

69. E. C. F. Forestier *et al.*, Developing a *Nicotiana benthamiana* transgenic platform for high-value diterpene production and candidate gene evaluation. *Plant Biotechnology Journal* **19**, 1614-1623 (2021).

70. A. Lossl *et al.*, Inducible trans-activation of plastid transgenes: Expression of the *R. eutropha* phb operon in transplastomic tobacco. *Plant and Cell Physiology* **46**, 1462-1471 (2005).

71. K. Bohmert, I. Balbo, A. Steinbuchel, G. Tischendorf, L. Willmitzer, Constitutive expression of the β-ketothiolase gene in transgenic plants: A major obstacle for obtaining polyhydroxybutyrate-producing plants. *Plant Physiology* **128**, 1282-1290 (2002).

72. L. Kourtz, K. Dillon, S. Daughtry, O. P. Peoples, K. D. Snell, Chemically inducible expression of the PHB biosynthetic pathway in *Arabidopsis*. *Transgenic Research* **16**, 759-769 (2007).

73. R. Ochoa-Fernandez *et al.*, Optogenetic control of gene expression in plants in the presence of ambient white light. *Nature Methods* **17**, 717-+ (2020).

74. L. Klewer, Y. W. Wu, Light-induced dimerization approaches to control cellular processes. *Chemistry-A European Journal* **25**, 12452-12463 (2019).

75. E. Gomes, J. Shorter, The molecular language of membraneless organelles. *Journal of Biological Chemistry* **294**, 7115-7127 (2019).

76. L. Wang *et al.*, A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell*, (2023).

77. S. Boeynaems *et al.*, Protein phase separation: A new phase in cell biology. *Trends in Cell Biology* **28**, 420-435 (2018).

78. W. Peeples, M. K. Rosen, Mechanistic dissection of increased enzymatic rate in a phase-separated compartment. *Nature Chemical Biology* **17**, 693-702 (2021).

79. L. J. Sweetlove, A. R. Fernie, The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nature Communications* **9**, (2018).

80. I. Wheeldon *et al.*, Substrate channelling as an approach to cascade reactions. *Nature Chemistry* **8**, 299-309 (2016).

81. H. Lee, W. C. DeLoache, J. E. Dueber, Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering* **14**, 242-251 (2012).

82. N. Smirnoff, Engineering of metabolic pathways using synthetic enzyme complexes. *Plant Physiology* **179**, 918-928 (2019).

83. I. Dahmani, K. Z. Qin, Y. J. Zhang, A. R. Fernie, The formation and function of plant metabolons. *The Plant Journal* **114**, 1080-1092 (2023).

84. M. Castellana *et al.*, Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology* **32**, 1011-+ (2014).

85. M. H. Abernathy, L. He, Y. J. Tang, Channeling in native microbial pathways: Implications and challenges for metabolic engineering. *Biotechnology Advances* **35**, 805-814 (2017).

86. M. de Jesus *et al.*, Tat proteins as novel thylakoid membrane anchors organize a biosynthetic pathway in chloroplasts and increase product yield 5-fold. *Metabolic Engineering* **44**, 108-116 (2017).

87. B. M. Long *et al.*, Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nature Communications* **9**, (2018).

88. L. C. Cheah *et al.*, Artificial self-assembling nanocompartment for organizing metabolic pathways in yeast. *ACS Synthetic Biology*, (2021).

89. Y. H. Lau, T. W. Giessen, W. J. Altenburg, P. A. Silver, Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nature Communications* **9**, (2018).

90. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).

91. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & Development* **33**, 1619-1634 (2019).

92. T. Mittag, R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell* **82**, 2201-2214 (2022).

93. A. Musacchio, On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal* **41**, e109952 (2022).

94. G. M. Ginell, A. S. Holehouse, in *Phase-Separated Biomolecular Condensates: Methods and Protocols,* H.-X. Zhou, J.-H. Spille, P. R. Banerjee, Eds. (Springer US, New York, NY, 2023), pp. 95-116.

95. A. Putnam, M. Cassani, J. Smith, G. Seydoux, A gel phase promotes condensation of liquid P granules in *Caenorhabditis elegans* embryos. *Nature Structural & Molecular Biology* **26**, 220-226 (2019).

96. T. Yamazaki *et al.*, Paraspeckles are constructed as block copolymer micelles. *The EMBO Journal* **40**, e107270 (2021).

97. D. De Sancho, Phase separation in amino acid mixtures is governed by composition. *Biophysical Journal* **121**, 4119-4127 (2022).

98. U. Rana, C. P. Brangwynne, A. Z. Panagiotopoulos, Phase separation vs aggregation behavior for model disordered proteins. *The Journal of Chemical Physics* **155**, 125101 (2021).

99. J. Solis-Miranda *et al.*, Stress-related biomolecular condensates in plants. *The Plant Cell* **35**, 3187-3204 (2023).

100. H. Glauninger, C. J. Wong Hickernell, J. A. M. Bard, D. A. Drummond, Stressful steps: Progress and challenges in understanding stress-induced mRNA condensation and accumulation in stress granules. *Molecular Cell* **82**, 2544-2556 (2022).

101. S. K. Powers *et al.*, Nucleo-cytoplasmic partitioning of ARF proteins controls auxin responses in *Arabidopsis thaliana*. *Molecular Cell* **76**, 177-190.e175 (2019).

102. B. G. O'Flynn, T. Mittag, A new phase for enzyme kinetics. *Nature Chemical Biology* **17**, 628-630 (2021).

103. S. Elbaum-Garfinkle *et al.*, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences* **112**, 7189-7194 (2015).

104. M. Guan *et al.*, Incorporation and assembly of a light-emitting enzymatic reaction into model protein condensates. *Biochemistry* **60**, 3137-3151 (2021).

105. Y. Wang *et al.*, Phase-separated multienzyme compartmentalization for terpene biosynthesis in a prokaryote. *Angewandte Chemie International Edition* **61**, e202203909 (2022).

106. M. V. Garabedian *et al.*, Protein condensate formation via controlled multimerization of intrinsically disordered sequences. *Biochemistry* **61**, 2470-2481 (2022).

107. M. Dzuricky, B. A. Rogers, A. Shahid, P. S. Cremer, A. Chilkoti, De novo engineering of intracellular condensates using artificial disordered proteins. *Nature Chemistry* **12**, 814-+ (2020).

108. A. Baruch Leshem *et al.*, Biomolecular condensates formed by designer minimalistic peptides. *Nature Communications* **14**, 421 (2023).

109. M. V. Garabedian *et al.*, Designer membraneless organelles sequester native factors for control of cell behavior. *Nature Chemical Biology* **17**, 998-+ (2021).

110. J. S. T. Ku, A. V. Y. Chen, E. I. Lan, Metabolic engineering design strategies for increasing acetyl-CoA flux. *Metabolites* **10**, (2020).

111. B. L. Fatland *et al.*, Molecular characterization of a heteromeric ATP-citrate lyase that generates cytosolic acetyl-coenzyme A in *Arabidopsis*. *Plant Physiology* **130**, 740-756 (2002).

112. M. Chypre, N. Zaidi, K. Smans, ATP-citrate lyase: A mini-review. *Biochemical and Biophysical Research Communications* **422**, 1-4 (2012).

113. J. Schwender, Y. Shachar-Hill, J. B. Ohlrogge, Mitochondrial metabolism in developing embryos of *Brassica napus*. *Journal of Biological Chemistry* **281**, 34040-34047 (2006).

114. R. Zhong *et al.*, Cytosolic acetyl-CoA generated by ATP-citrate lyase is essential for acetylation of cell wall polysaccharides. *Plant and Cell Physiology* **61**, 64-75 (2020).

115. S. Xing, Y. Poirier, The protein acetylome and the regulation of metabolism. *Trends in Plant Science* **17**, 423-430 (2012).

116. M. Kanehisa *et al.*, From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**, D354-D357 (2006).

117. P. D. Karp *et al.*, Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research* **33**, 6083-6089 (2005).

118. C. J. Foster, L. Wang, H. V. Dinh, P. F. Suthers, C. D. Maranas, Building kinetic models for metabolic engineering. *Current Opinion in Biotechnology* **67**, 35-41 (2021).

119. S. Dash *et al.*, Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnology for Biofuels* **10**, 108 (2017).

120. L. Miskovic *et al.*, A design–build–test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant *S. cerevisiae* and improves predictive capabilities of large-scale kinetic models. *Biotechnology for Biofuels* **10**, 166 (2017).

121. M. G. Poolman, H. E. Assmus, D. A. Fell, Applications of metabolic modelling to plant metabolism. *Journal of Experimental Botany* **55**, 1177-1186 (2004).

122. D. Davidi *et al.*, Global characterization of *in vivo* enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proceedings of the National Academy of Sciences* **113**, 3401-3406 (2016).

123. S. Gopalakrishnan, S. Dash, C. Maranas, K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabolic Engineering* **61**, 197-205 (2020).

124. H. Yu, H. Deng, J. He, J. D. Keasling, X. Luo, UniKP: a unified framework for the prediction of enzyme kinetic parameters. *Nature Communications* **14**, 8211 (2023).

125. A. Sahu, M.-A. Blätke, J. J. Szymański, N. Töpfer, Advances in flux balance analysis by integrating machine learning and mechanism-based models. *Computational and Structural Biotechnology Journal* **19**, 4626-4640 (2021).

126. D. Segrè, D. Vitkup, G. M. Church, Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences* **99**, 15112-15117 (2002).

127. H. A. Herrmann, B. C. Dyson, L. Vass, G. N. Johnson, J. M. Schwartz, Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *Npj Systems Biology and Applications* **5**, (2019).

128. K. J. van Wijk, Plastid proteomics. *Plant Physiology and Biochemistry* **42**, 963-977 (2004).

129. E. Martinoia, M. Maeshima, H. E. Neuhaus, Vacuolar transporters and their essential role in plant metabolism. *Journal of Experimental Botany* **58**, 83-102 (2007).

130. C. Pozo *et al.*, Gobal optimization of hybrid kinetic/FBA models via outer-approximation. *Computers & Chemical Engineering* **72**, 325-333 (2015).

131. E. Bogart, C. R. Myers, Multiscale metabolic modeling of C4 plants: Connecting nonlinear genome-scale models to leaf-scale metabolism in developing maize leaves. *PLOS One* **11**, e0151722 (2016).

132. A. P. Burgard, P. Pharkya, C. D. Maranas, Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering* **84**, 647-657 (2003).

133. S. S. Fong *et al.*, *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnology and Bioengineering* **91**, 643-648 (2005).

134. H. Alper, Y.-S. Jin, J. F. Moxley, G. Stephanopoulos, Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metabolic Engineering* **7**, 155-164 (2005).

135. J. M. Lee, E. P. Gianchandani, J. A. Papin, Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics* **7**, 140-150 (2006).

136. B. Dugdale *et al.*, In plant activation: An inducible, hyperexpression platform for recombinant protein production in plants. *The Plant Cell* **25**, 2429-2443 (2013).

137. M. Shimamura, *Marchantia polymorpha*: Taxonomy, phylogeny and morphology of a model system. *Plant and Cell Physiology* **57**, 230-256 (2016).

138. J. L. Bowman, T. Araki, T. Kohchi, *Marchantia*: Past, present and future. *Plant and Cell Physiology* **57**, 205-209 (2016).

139. N. Cannell *et al.*, Multiple metabolic innovations and losses are associated with major transitions in land plant evolution. *Current Biology* **30**, 1783-+ (2020).

140. N. Cannell, University of Oxford, (2021).

141. S. Shameer, K. Baghalian, C. Y. M. Cheung, R. G. Ratcliffe, L. J. Sweetlove, Computational analysis of the productivity potential of CAM. *Nature Plants* **4**, 165-171 (2018).

142. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, (2015).

143. J. L. Bowman *et al.*, Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287-304.e215 (2017).

144. K. Happ, B. Classen, Arabinogalactan-proteins from the liverwort *Marchantia polymorpha* L., a member of a basal land plant lineage, are structurally different to those of Angiosperms. *Plants (Basel)* **8**, (2019).

145. V. M. Dembitsky, T. Rezanka, I. A. Bychek, O. M. Afonina, Polar lipid and fatty acid composition of some bryophytes. *Phytochemistry* **33**, 1009-1014 (1993).

146. Y. Shinmen, K. Katoh, S. Shimizu, S. Jareonkitmongkol, H. Yamada, Production of arachidonic acid and eicosapentaenoic acids by *Marchantia polymorpha* in cell culture. *Phytochemistry* **30**, 3255-3260 (1991).

147. R. J. Thomas, F. J. Behringer, C. S. Lombard, J. J. Sparkowski, Effects of auxin on wall polysaccharide composition and enzyme-activity during extension-growth of *Pellia* (*Bryophyta*). *Physiologia Plantarum* **60**, 502-506 (1984).

148. E. Koide *et al.*, Regulation of photosynthetic carbohydrate metabolism by a Raf-like kinase in the liverwort *Marchantia polymorpha*. *Plant and Cell Physiology* **61**, 631-643 (2020).

149. P. Carella *et al.*, Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Current Biology* **29**, 2282-2294.e2285 (2019).

150. C. Y. M. Cheung, M. G. Poolman, D. A. Fell, R. G. Ratcliffe, L. J. Sweetlove, A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C-3 and crassulacean acid metabolism Leaves. *Plant Physiology* **165**, 917-929 (2014).

151. K. Katoh, Photosynthesis and photoassimilation of glucose during photomixotrophic growth of *Marchantia polymorpha* cells in suspension-culture. *Physiologia Plantarum* **57**, 67-74 (1983).

152. R. J. Thomas, C. S. Lombard, Aphid infestation and its effects on translocation in *Polytrichum commune*. *Bryologist* **94**, 1-4 (1991).

153. O. P. Peoples, A. J. Sinskey, Poly-beta-hydroxybutyrate biosynthesis in *Alcaligenes eutrophus* H16 - Characterization of the genes encoding β-ketothiolase and acetoacetyl-CoA reductase. *Journal of Biological Chemistry* **264**, 15293-15297 (1989).

154. X. Guan, B. J. Nikolau, AAE13 encodes a dual-localized malonyl-CoA synthetase that is crucial for mitochondrial fatty acid biosynthesis. *The Plant Journal* **85**, 581-593 (2016).

155. Y. Sasaki, Y. Nagano, Plant acetyl-CoA carboxylase: Structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience Biotechnology and Biochemistry* **68**, 1175-1184 (2004).

156. H. Chen, H. U. Kim, H. Weng, J. Browse, Malonyl-CoA synthetase, encoded by ACYL ACTIVATING ENZYME13, is essential for growth and development of *Arabidopsis*. *Plant Cell* **23**, 2247-2262 (2011).

157. S. Shameer, R. G. Ratcliffe, L. J. Sweetlove, Leaf energy balance requires mitochondrial respiration and export of chloroplast NADPH in the light. *Plant Physiology* **180**, 1947-1961 (2019).

158. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *Bmc Systems Biology* **7**, (2013).

159. N. E. Lewis, H. Nagarajan, B. O. Palsson, Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nature Reviews Microbiology* **10**, 291-305 (2012).

160. S. Gudmundsson, I. Thiele, Computationally efficient flux variability analysis. *Bmc Bioinformatics* **11**, (2010).

161. W. Megchelenbrink, M. Huynen, E. Marchiori, optGpSampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *Plos One* **9**, (2014).

162. J. Ray, A. Pinar, C. Seshadhri, in *Algorithms and Models for the Web Graph,* A. Bonato, J. Janssen, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), pp. 153-164.

163. M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence Diagnosis and Output Analysis for MCMC. *R news* **6**, 7-11 (2006).

164. W. H. Kruskal, W. A. Wallis, Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**, 583-621 (1952).

165. T. Kluyver *et al.*, in *IOS Press*. (2016), pp. 87-90.

166. H. J. Keselman, J. C. Rogan, The Tukey multiple comparison test: 1953–1976. *Psychological Bulletin* **84**, 1050-1056 (1977).

167. H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**, 50-60 (1947).

168. P. Virtanen *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).

169. M. Waskom *et al.*, *seaborn: Statistical data visualization*. (2020), pp. ascl:2012.2015.

170. J. D. Hunter, Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**, 90-95 (2007).

171. F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

172. A. Sándor, University of Oxford, Oxford (2022).

173. B. T. Bajar *et al.*, Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports* **6**, 20889 (2016).

174. M. Karimi, D. Inzé, A. Depicker, GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* **7**, 193-195 (2002).

175. P. Puigbo, E. Guzman, A. Romeu, S. Garcia-Vallve, OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research* **35**, W126-W131 (2007).

176. A. L. Szymczak-Workman, K. M. Vignali, D. A. A. Vignali, Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harbor protocols* **2012**, 199-204 (2012).

177. M. L. L. Donnelly *et al.*, The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *Journal of General Virology* **82**, 1027-1041 (2001).

178. Y. C. Wang, F. Wang, R. Y. Wang, P. Zhao, Q. Y. Xia, 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Scientific Reports* **5**, (2015).

179. B. Zhang *et al.*, Coordinated protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide. *Plant Biotechnology Journal* **15**, 718-728 (2017).

180. P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102-+ (2000).

181. B. Frei *et al.*, Purification and functional characterization of the vacuolar malate transporter tDT from *Arabidopsis*. *Journal of Biological Chemistry* **293**, 4180-4190 (2018).

182. B. S. Schuster *et al.*, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences* **117**, 11421-11431 (2020).

183. K. E. Thompson, C. J. Bashor, W. A. Lim, A. E. Keating, SYNZIP protein interaction toolbox: *in vitro* and *in vivo* specifications of heterospecific coiled-coil interaction domains. *ACS Synthetic Biology* **1**, 118-129 (2012).

184. N. Sugimoto, P. Engelgau, A. D. Jones, J. Song, R. Beaudry, Citramalate synthase yields a biosynthetic pathway for isoleucine and straight- and branched-chain ester formation in ripening apple fruit. *Proceedings of the National Academy of Sciences* **118**, e2009988118 (2021).

185. N. Sharma, M. X. Anleu Gil, D. Wengier, A quick and easy method for making competent *Escherichia coli* cells for transformation using rubidium chloride. *Bio-101* **7**, e2590 (2017).

186. C. Koncz, J. Schell, The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Molecular and General Genetics MGG* **204**, 383-396 (1986).

187. F. Jay, F. Brioudes, O. Voinnet, A contemporary reassessment of the enhanced transient expression system based on the tombusviral silencing suppressor protein P19. *The Plant Journal* **113**, 186-204 (2023).

188. S. Porebski, L. G. Bailey, B. R. Baum, Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**, 8-15 (1997).

189. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).

190. S. Bolte, F. P. CordeliÈRes, A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy* **224**, 213-232 (2006).

191. B. S. Schuster *et al.*, Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nature Communications* **9**, (2018).

192. J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A. R. Fernie, Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nature Protocols* **1**, 387-396 (2006).

193. K. Kocharin, V. Siewers, J. Nielsen, Improved polyhydroxybutyrate production by *Saccharomyces cerevisiae* through the use of the phosphoketolase pathway. *Biotechnology and Bioengineering* **110**, 2216-2224 (2013).

194. R. Carlson, D. Fell, F. Srienc, Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering* **79**, 121-134 (2002).

195. J. M. Rohwer, Kinetic modelling of plant metabolic pathways. *Journal of Experimental Botany* **63**, 2275-2292 (2012).

196. T. B. Moreira, J. M. Lima, G. C. Coca, T. C. R. Williams, Insights into the spatial and temporal organisation of plant metabolism from network flux analysis. *Theoretical and Experimental Plant Physiology* **31**, 215-226 (2019).

197. K. A. Curran, N. C. Crook, H. S. Alper, in *Microbial Metabolic Engineering: Methods and Protocols,* Q. Cheng, Ed. (Springer New York, New York, NY, 2012), pp. 197-216.

198. S. Hashemi, Z. Razaghi-Moghadam, Z. Nikoloski, Maximizing multi-reaction dependencies provides more accurate and precise predictions of intracellular fluxes than the principle of parsimony. *PLOS Computational Biology* **19**, e1011489 (2023).

199. L. J. Sweetlove, K. F. M. Beard, A. Nunes-Nesi, A. R. Fernie, R. G. Ratcliffe, Not just a circle: flux modes in the plant TCA cycle. *Trends in Plant Science* **15**, 462-470 (2010).

200. Y. Gibon *et al.*, Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. *Genome Biology* **7**, R76 (2006).

201. K. Tilbrook, L. Gebbie, P. M. Schenk, Y. Poirier, S. M. Brumbley, Peroxisomal polyhydroxyalkanoate biosynthesis is a promising strategy for bioplastic production in high biomass crops. *Plant Biotechnology Journal* **9**, 958-969 (2011).

202. S. Weise. (2013), vol. 2024.

203. J. Ke *et al.*, The role of pyruvate dehydrogenase and acetyl-coenzyme A synthetase in fatty acid synthesis in developing *Arabidopsis* seeds. *Plant Physiology* **123**, 497-508 (2000).

204. R. Mahadevan, C. H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering* **5**, 264-276 (2003).

205. A. Etienne, M. Genard, P. Lobit, D. Mbeguie-A-Mbeguie, C. Bugaud, What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. *Journal of Experimental Botany* **64**, 1451-1469 (2013).

206. I. Pracharoenwattana *et al.*, *Arabidopsis* has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *The Plant Journal* **62**, 785-795 (2010).

207. M. A. Hurth *et al.*, Impaired pH homeostasis in *Arabidopsis* lacking the vacuolar dicarboxylate transporter and analysis of carboxylic acid transport across the tonoplast. *Plant Physiology* **137**, 901-910 (2005).

208. X.-Y. Huang, C.-K. Wang, Y.-W. Zhao, C.-H. Sun, D.-G. Hu, Mechanisms and regulation of organic acid accumulation in plant vacuoles. *Horticulture Research* **8**, (2021).

209. Y. B. Li *et al.*, Evolution of tonoplast P-ATPase transporters involved in vacuolar acidification. *New Phytologist* **211**, 1092-1107 (2016).

210. B. Ma *et al.*, A Ma10 gene encoding P-type ATPase is involved in fruit organic acid accumulation in apple. *Plant Biotechnology Journal* **17**, 674-686 (2019).

211. I. Appelhagen *et al.*, TRANSPARENT TESTA 13 is a tonoplast P-3A-ATPase required for vacuolar deposition of proanthocyanidins in *Arabidopsis thaliana* seeds. *The Plant Journal* **82**, 840-849 (2015).

212. C. P. Lee *et al.*, The versatility of plant organic acid metabolism in leaves is underpinned by mitochondrial malate-citrate exchange. *The Plant cell*, (2021).

213. G. A. Bazilevsky *et al.*, ATP-citrate lyase multimerization is required for coenzyme-A substrate binding and catalysis. *Journal of Biological Chemistry* **294**, 7259-7268 (2019).

214. Q. Zhang, W. Zeng, S. Xu, J. Zhou, Metabolism and strategies for enhanced supply of acetyl-CoA in *Saccharomyces cerevisiae*. *Bioresource Technology* **342**, 125978 (2021).

215. T. Yu *et al.*, Metabolic reconfiguration enables synthetic reductive metabolism in yeast. *Nature Metabolism* **4**, 1551-1559 (2022).

216. X. Feng, J. Lian, H. Zhao, Metabolic engineering of *Saccharomyces cerevisiae* to improve 1-hexadecanol production. *Metabolic Engineering* **27**, 10-19 (2015).

217. J. Lian, T. Si, N. U. Nair, H. Zhao, Design and construction of acetyl-CoA overproducing *Saccharomyces cerevisiae* strains. *Metabolic Engineering* **24**, 139-149 (2014).

218. A. Ghosh *et al.*, 13C metabolic flux analysis for systematic metabolic engineering of *S. cerevisiae* for overproduction of fatty acids. *Frontiers in Bioengineering and Biotechnology* **4**, (2016).

219. X. Tang, H. Feng, W. N. Chen, Metabolic engineering for enhanced fatty acids synthesis in *Saccharomyces cerevisiae*. *Metabolic Engineering* **16**, 95-102 (2013).

220. D. H. Kim *et al.*, Metabolic engineering of *Saccharomyces cerevisiae* by using the CRISPR-Cas9 system for enhanced fatty acid production. *Process Biochemistry* **73**, 23-28 (2018).

221. C. G. Andersen, L. Bavnhøj, B. P. Pedersen, May the proton motive force be with you: A plant transporter review. *Current Opinion in Structural Biology* **79**, 102535 (2023).

222. M. Tamoi, M. Nagaoka, Y. Miyagawa, S. Shigeoka, Contribution of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase to the photosynthetic rate and carbon flow in the Calvin cycle in transgenic plants. *Plant and Cell Physiology* **47**, 380-390 (2006).

223. E. Gruchattka, O. Kayser, *In vivo* validation of *in silico* predicted metabolic engineering strategies in yeast: Disruption of alpha-ketoglutarate dehydrogenase and expression of ATP-citrate lyase for terpenoid production. *PLOS One* **10**, (2015).

224. I. Finkemeier *et al.*, Transcriptomic analysis of the role of carboxylic acids in metabolite signaling in *Arabidopsis* leaves. *Plant Physiology* **162**, 239-253 (2013).

225. A. U. Igamberdiev, A. T. Eprintsev, Organic acids: The pools of fixed carbon involved in redox regulation and energy balance in higher plants. *Frontiers in Plant Science* **7**, (2016).

226. S. Tsuboyama, Y. Kodama, AgarTrap protocols on your benchtop: Simple methods for *Agrobacterium*-mediated genetic transformation of the liverwort *Marchantia polymorpha*. *Plant Biotechnology* **35**, 93-99 (2018).

227. S. Tsuboyama-Tanaka, Y. Kodama, AgarTrap-mediated genetic transformation using intact gemmae/gemmalings of the liverwort *Marchantia polymorpha* L. *J Plant Res* **128**, 337-344 (2015).

228. F. Althoff *et al.*, Comparison of the MpEF1α and CaMV35 promoters for application in *Marchantia polymorpha* overexpression studies. *Transgenic Research* **23**, 235-244 (2014).

229. J. Zuo, Q. W. Niu, N. H. Chua, Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant Journal* **24**, 265-273 (2000).

230. E. Flores-Sandoval, T. Dierschke, T. J. Fisher, J. L. Bowman, Efficient and inducible use of artificial microRNAs in *Marchantia polymorpha*. *Plant and Cell Physiology* **57**, 281-290 (2016).

231. K. Ishizaki, M. Nonomura, H. Kato, K. T. Yamato, T. Kohchi, Visualization of auxin-mediated transcriptional activation using a common auxin-responsive reporter system in the liverwort *Marchantia polymorpha*. *Journal of Plant Research* **125**, 643-651 (2012).

232. B. Larsen *et al.*, Highlighter: An optogenetic system for high-resolution gene expression control in plants. *PLOS Biology* **21**, e3002303 (2023).

233. H. Shikata, P. Denninger, Plant optogenetics: Applications and perspectives. *Current Opinion in Plant Biology* **68**, 102256 (2022).

234. A. N. Binns, M. F. Thomashow, Cell biology of *Agrobacterium* infection and transformation of plants. *Annual Review of Microbiology* **42**, 575-606 (1988).

235. K. Ishizaki, R. Nishihama, K. T. Yamato, T. Kohchi, Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant and Cell Physiology* **57**, 262-270 (2016).

236. A. Wilmink, J. J. M. Dons, Selective agents and marker genes for use in transformation of monocotyledonous plants. *Plant Molecular Biology Reporter* **11**, 165-185 (1993).

237. K. Muller *et al.*, A red light-controlled synthetic gene expression switch for plant systems. *Molecular Biosystems* **10**, 1679-1688 (2014).

238. A. S. Shifera, J. A. Hardin, Factors modulating expression of *Renilla* luciferase from control plasmids used in luciferase reporter gene assays. *Analytical Biochemistry* **396**, 167-172 (2010).

239. D. J. Heyes *et al.*, Light-induced structural changes in a full-length cyanobacterial phytochrome probed by time-resolved X-ray scattering. *Communications Biology* **2**, 1 (2019).

240. B. D. Zoltowski, A. I. Nash, K. H. Gardner, Variations in protein–flavin hydrogen bonding in a light, oxygen, voltage domain produce non-Arrhenius kinetics of adduct decay. *Biochemistry* **50**, 8771-8779 (2011).

241. J. E. Solly, N. J. Cunniffe, C. J. Harrison, Regional growth rate differences specified by apical notch activities regulate liverwort thallus shape. *Current Biology* **27**, 16-26 (2017).

242. K. Salin *et al.*, Differences in mitochondrial efficiency explain individual variation in growth performance. *Proceedings of the Royal Society B: Biological Sciences* **286**, 20191466 (2019).

243. B. A. Sherf, S. L. Navarro, R. R. Hannah, K. V. Wood. (1996).

244. M. Delmans, B. Pollak, J. Haseloff, MarpoDB: An open registry for *Marchantia polymorpha* genetic parts. *Plant and Cell Physiology* **58**, e5 (2017).

245. J. Westermann, E. Koebke, R. Lentz, M. Hülskamp, A. Boisson-Dernier, A comprehensive toolkit for quick and easy visualization of marker proteins, protein-protein interactions and cell morphology in *Marchantia polymorpha*. *Frontiers in Plant Science* **11**, 569194 (2020).

246. L. B. Motta-Mena *et al.*, An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nature Chemical Biology* **10**, 196-202 (2014).

247. M.-C. Cheng, P. K. Kathare, I. Paik, E. Huq, Phytochrome signaling networks. *Annual Review of Plant Biology* **72**, 217-244 (2021).

248. K. Inoue *et al.*, Phytochrome signaling is mediated by PHYTOCHROME INTERACTING FACTOR in the liverwort *Marchantia polymorpha*. *Plant Cell* **28**, 1406-1421 (2016).

249. R. Nishihama *et al.*, Phytochrome-mediated regulation of cell division and growth during regeneration and sporeling development in the liverwort *Marchantia polymorpha*. *Journal of Plant Research* **128**, 407-421 (2015).

250. S. Streubel *et al.*, Meristem dormancy in *Marchantia polymorpha* is regulated by a liverwort-specific miRNA and a clade III *SPL* gene. *Current Biology* **33**, 660-674.e664 (2023).

251. R. Nishihama, S. Ishida, H. Urawa, Y. Kamei, T. Kohchi, Conditional gene expression/deletion systems for *Marchantia polymorpha* using its own heat-shock promoter and Cre/loxP-mediated site-specific recombination. *Plant and Cell Physiology* **57**, 271-280 (2016).

252. Z. Liu *et al.*, Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports* **7**, (2017).

253. J. H. Kim *et al.*, High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLOS One* **6**, (2011).

254. L. Wang *et al.*, The maturation and aging trajectory of *Marchantia polymorpha* at single-cell resolution. *Developmental Cell* **58**, 1429-1444.e1426 (2023).

255. D. Jendrossek, O. Selchow, M. Hoppert, Poly(3-hydroxybutyrate) granules at the early stages of formation are localized close to the cytoplasmic membrane in *Caryophanon latum*. *Applied and Environmental Microbiology* **73**, 586-593 (2007).

256. L. G. Izzo, B. Hay Mele, L. Vitale, E. Vitale, C. Arena, The role of monochromatic red and blue light in tomato early photomorphogenesis and photosynthetic traits. *Environmental and Experimental Botany* **179**, 104195 (2020).

257. E. Kaiser *et al.*, Adding blue to red supplemental light increases biomass and yield of greenhouse-grown tomatoes, but only to an optimum. *Frontiers in Plant Science* **9**, 2002 (2018).

258. K. H. G. Verschueren *et al.*, Structure of ATP citrate lyase and the origin of citrate synthase in the Krebs cycle. *Nature* **568**, 571-+ (2019).

259. R. Lin *et al.*, Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth. *Molecular Cell* **51**, 506-518 (2013).

260. T. Shimada, R. Nakano, V. Shulaev, A. Sadka, E. Blumwald, Vacuolar citrate/H+ symporter of citrus juice cells. *Planta* **224**, 472-480 (2006).

261. J. P. Zubimendi *et al.*, The complex allosteric and redox regulation of the fumarate hydratase and malate dehydratase reactions of *Arabidopsis thaliana* Fumarase 1 and 2 gives clues for understanding the massive accumulation of fumarate. *The FEBS Journal* **285**, 2205-2224 (2018).

262. J. Ye *et al.*, An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *The Plant Cell* **29**, 2249-2268 (2017).

263. K. Sakoda *et al.*, Higher stomatal density improves photosynthetic induction and biomass production in *Arabidopsis* under fluctuating light. *Frontiers in Plant Science* **11**, 589603 (2020).

264. B. Xu *et al.*, The land plant-specific MIXTA-MYB lineage is implicated in the early evolution of the plant cuticle and the colonization of land. *New Phytologist* **229**, 2324-2338 (2021).

265. D. M. Eklund *et al.*, An evolutionarily conserved abscisic acid signaling pathway regulates dormancy in the liverwort *Marchantia polymorpha*. *Current Biology* **28**, 3691-3699.e3693 (2018).

266. K. Tilbrook *et al.*, Reduced peroxisomal citrate synthase activity increases substrate availability for polyhydroxyalkanoate biosynthesis in plant peroxisomes. *Plant Biotechnology Journal* **12**, 1044-1052 (2014).

267. C. K. Holland, J. M. Jez, Reaction mechanism of prephenate dehydrogenase from the alternative tyrosine biosynthesis pathway in plants. *ChemBioChem* **19**, 1132-1136 (2018).

268. O. R. A. Corea *et al.*, Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *Journal of Biological Chemistry* **287**, 11446-11459 (2012).

269. J. Peek, D. Christendat, The shikimate dehydrogenase family: functional diversity within a conserved structural and mechanistic framework. *Archives of Biochemistry and Biophysics* **566**, 85-99 (2015).

270. J. Ping, L. Wang, Z. Qin, Z. Zhou, J. Zhou, Synergetic engineering of *Escherichia coli* for efficient production of l-tyrosine. *Synthetic and Systems Biotechnology* **8**, 724-731 (2023).

271. M. Trovato, D. Funck, G. Forlani, S. Okumoto, R. Amir, Amino acids in plants: Regulation and functions in development and stress defense. *Frontiers in Plant Science* **12**, 772810 (2021).

272. T. Hirota, M. Izumi, S. Wada, A. Makino, H. Ishida, Vacuolar protein degradation via autophagy provides substrates to amino acid catabolic pathways as an adaptive response to sugar starvation in *Arabidopsis thaliana*. *Plant and Cell Physiology* **59**, 1363-1376 (2018).

273. G. Seydoux, The P granules of *C. elegans*: A genetic model for the study of RNA-protein condensates. *Journal of Molecular Biology* **430**, 4702-4710 (2018).

274. N. C. Shaner *et al.*, Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology* **22**, 1567-1572 (2004).

275. R. J. Emenecker, A. S. Holehouse, L. C. Strader, Biological phase separation and biomolecular condensates in plants. *Annual Review of Plant Biology* **72**, 17-46 (2021).

276. G. P. Doherty, K. Bailey, P. J. Lewis, Stage-specific fluorescence intensity of GFP and mCherry during sporulation in *Bacillus subtilis*. *BMC Research Notes* **3**, 303 (2010).

277. E. M. M. Manders, F. J. Verbeek, J. A. Aten, Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy* **169**, 375-382 (1993).

278. S. Fujimoto, M. Ito, S. Matsunaga, K. Fukui, An upper limit of the ratio of DNA volume to nuclear volume exists in plants. *Genes & Genetic Systems* **80**, 345-350 (2005).

279. D. Tolleter *et al.*, The *Arabidopsis* leaf quantitative atlas: a cellular and subcellular mapping through unified data integration. *Quantitative Plant Biology* **5**, e2 (2024).

280. S. Alberti, A. Gladfelter, T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419-434 (2019).

281. C. De Los Santos, C. W. Chang, M. A. Mycek, R. A. Cardullo, FRAP, FLIM, and FRET: Detection and analysis of cellular dynamics on a molecular scale using fluorescence microscopy. *Molecular Reproduction and Development* **82**, 587-604 (2015).

282. M. Kang, C. A. Day, A. K. Kenworthy, E. DiBenedetto, Simplified equation to extract diffusion coefficients from confocal FRAP data. *Traffic* **13**, 1589-1600 (2012).

283. F. Madeira *et al.*, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research* **50**, W276-W279 (2022).

284. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

285. A. Jose Juan Almagro *et al.*, Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* **2**, e201900429 (2019).

286. J. Shen *et al.*, Organelle pH in the *Arabidopsis* endomembrane system. *Molecular Plant* **6**, 1419-1437 (2013).

287. M. Kyoung, S. J. Russell, C. L. Kohnhorst, N. N. Esemoto, S. An, Dynamic architecture of the purinosome involved in human *de novo* purine biosynthesis. *Biochemistry* **54**, 870-880 (2015).

288. R. E. Lawrence *et al.*, A nutrient-induced affinity switch controls mTORC1 activation by its Rag GTPase–Ragulator lysosomal scaffold. *Nature Cell Biology* **20**, 1052-1063 (2018).

289. F. Mueller, D. Mazza, T. J. Stasevich, J. G. McNally, FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know? *Current Opinion in Cell Biology* **22**, 403-411 (2010).

290. N. O. Taylor, M.-T. Wei, H. A. Stone, C. P. Brangwynne, Quantifying dynamics in phase-separated condensates using fluorescence recovery after photobleaching. *Biophysical journal* **117**, 1285-1300 (2019).

291. D. T. McSwiggen *et al.*, Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *Elife* **8**, (2019).

292. J. L. Bowman *et al.*, The renaissance and enlightenment of *Marchantia* as a model system. *Plant Cell* **34**, 3512-3542 (2022).

293. S. Nagaya, M. Takemura, K. Ohyama, Endogenous promoter, 5 '-UTR and transcriptional terminator enhance transient gene expression in a liverwort, *Marchantia polymorpha* L. *Plant Biotechnology* **28**, 493-496 (2011).

294. M. Takemura, S. Nagaya, H. Kida, Effect of the 5'-untranslated region on the expression of transgene in liverwort, *Marchantia polymorpha*. *Plant Biotechnology* **32**, 317-321 (2015).

295. E. Ávila, A. Herrera, W. Tezara, Contribution of stem CO_2 fixation to whole-plant carbon balance in nonsucculent species. *Photosynthetica* **52**, 3-15 (2014).

296. T. Elmayan, M. Tepfer, Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter. *Transgenic Research* **4**, 388-396 (1995).

297. E. Frangedakis *et al.*, Construction of DNA tools for hyperexpression in *Marchantia* chloroplasts. *ACS Synthetic Biology* **10**, 1651-1666 (2021).

298. Y. Lu, F. F. Eiriksson, M. Thorsteinsdóttir, H. T. Simonsen, Valuable fatty acids in bryophytes—production, biosynthesis, analysis and applications. *Plants*. 2019 (10.3390/plants8110524).

299. Y. Asakawa, in *Progress in the Chemistry of Organic Natural Products,* Y. Asakawa, Ed. (Springer Vienna, Vienna, 1995), pp. 1-562.

300. B. Singh, R. A. Sharma, Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **5**, 129-151 (2015).

301. B. L. Fatland, B. J. Nikolau, E. S. Wurtele, Reverse genetic characterization of cytosolic acetyl-CoA generation by ATP-citrate lyase in *Arabidopsis*. *Plant Cell* **17**, 182-203 (2005).

302. A. Steinbuchel *et al.*, Considerations on the structure and biochemistry of bacterial polyhydroxyalkanoic acid inclusions. *Canadian Journal of Microbiology* **41 Suppl 1**, 94-105 (1995).

303. A. Lechner, E. Brunk, J. D. Keasling, The need for integrated approaches in metabolic engineering. *Cold Spring Harbor Perspectives in Biology* **8**, (2016).

304. K. E. J. Tyo, C. R. Fischer, F. Simeon, G. Stephanopoulos, Analysis of polyhydroxybutyrate flux limitations by systematic genetic and metabolic perturbations. *Metabolic Engineering* **12**, 187-195 (2010).

305. F. Wang, S. Y. Lee, Production of poly(3-hydroxybutyrate) by fed-batch culture of filamentation-suppressed recombinant *Escherichia coli*. *Applied and Environmental Microbiology* **63**, 4765-4769 (1997).

306. R. J. van Wegen, S.-Y. Lee, A. P. J. Middelberg, Metabolic and kinetic analysis of poly(3-hydroxybutyrate) production by recombinant *Escherichia coli*. *Biotechnology and Bioengineering* **74**, 70-81 (2001).

307. T. Yokoo, K. Matsumoto, T. Ooba, K. Morimoto, S. Taguchi, Enhanced poly(3-hydroxybutyrate) production in transgenic tobacco BY-2 cells using engineered acetoacetyl-CoA reductase. *Bioscience Biotechnology and Biochemistry* **79**, 986-988 (2015).

308. P. Borrill, Blurring the boundaries between cereal crops and model plants. *New Phytologist* **228**, 1721-1727 (2020).

309. L. Sethi, K. Kumari, N. Dey, Engineering of plants for efficient production of therapeutics. *Molecular Biotechnology* **63**, 1125-1137 (2021).

310. P. Delhon, A. Gojon, P. Tillard, L. Passama, Diurnal regulation of NO_3- uptake in soybean plants. *Journal of Experimental Botany* **46**, 1585-1594 (1995).

311. J. H. Macduff, A. K. Bakken, Diurnal variation in uptake and xylem contents of inorganic and assimilated N under continuous and interrupted N supply to *Phleum pratense* and *Festuca pratensis*. *Journal of Experimental Botany* **54**, 431-444 (2003).

312. Y. Gibon *et al.*, Adjustment of diurnal starch turnover to short days: Depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period. *The Plant Journal* **40**, 332-332 (2004).

313. J. L. Carter, K. G. Romine, The effects of long and short photoperiods on the rate of growth and gemmae cup production in *Marchantia polymorpha* L. *Transactions of the Kansas Academy of Science (1903-)* **72**, 98-107 (1969).

314. F. Ma, L. J. Jazmin, J. D. Young, D. K. Allen, Isotopically nonstationary C-13 flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 16967-16972 (2014).

315. W. R. Scheible, A. Krapp, M. Stitt, Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant Cell and Environment* **23**, 1155-1167 (2000).

Supplementary Information

Table S2.1: Constraints used for the generation of a stoichiometric model of *M. polymorpha* central metabolism. Upper_bound: maximum allowable flux value; lower_bound: minimum allowable flux value.

Constraint	Value (fluxes in units of µmol m^{-2} s^{-1})	Reasoning
Light input during the night	Upper_bound = 0	This allows for concurrent modelling of light and dark phases with a 12 h:12 h photoperiod
Nitrate uptake	Constrained to 3:2 day:night ratio	Based on measurements in several angiosperms (*310, 311*).
Biomass	Upper_bound = 0.238, constrained 3:1 between day and night	The output of a *Marchantia*-specific biomass equation was constrained 3:1 between day and night (*312*) up to a maximum flux of 0.238 µmol biomass m^{-2} s^{-1}, an experimentally determined growth rate for *M. polymorpha* (*313*).
Maintenance	Flux through ATP hydrolase constrained to 1.895; equal between day and night phases	Energetic costs of NGAM were modelled using constraints on ATP hydrolysis and NADPH oxidase reaction fluxes. This includes a constraint on proportions of flux through ATP hydrolysis and NADPH oxidase (constrained to a 3:1 ratio based on (*150*)), a value for NGAM estimated using data from (*151*), NGAM costs assumed to be equal in the day and the night, and the flux through NADPH oxidase spread evenly across plastid, cytosol, and mitochondrion.
Rubisco carboxylase/oxygenase activity	Constrained in a 3:1 ratio	Based on experimentally measured fluxes in *A. thaliana (314)*.
Accumulation of amino acids	Allowed only in the day	Based on measurements in tobacco leaves (*315*).
Photon uptake	Upper_bound = 174	Light availability associated with experimental growth rate of 0.238 µmol biomass m^{-2} s^{-1} , the constraint used on biomass accumulation in

		this model (*313*).
CO_2 assimilation	CO_2 fixation during the night is not permitted, and assimilation constrained to 1.16 µmol CO_2 m^{-2} s^{-1}	The FBA procedure will lead to a maximally optimised model – if not constrained this results in the model re-fixing all CO_2 at night which is unrealistic for C_3 organisms. Constrained to calculated carbon assimilation rate when 174 µmol photons m^{-2} s^{-1} available (*313*).
Palmitate uptake	Upper and lower_bound = 0	A constraint on the uptake of palmitate was added based on the observation that the model was taking up palmitate from the environment rather than synthesizing it *de novo*.
Diel transfer of malate	Upper and lower_bound = 0	Experimentally derived (Corinna Hartinger, unpublished)
Diel transfer of fumarate	Upper and lower_bound = 0	Experimentally derived (Corinna Hartinger, unpublished)
Diel transfer of citrate	Upper_bound = -0.012057; lower_bound = -0.013223	Experimentally derived (Corinna Hartinger, unpublished); the experimental findings were converted to the units of the model and used as a constraint on the diel flux of citrate.
Amount of assimilated carbon exported to the rhizoids	4.3 % of assimilated carbon; 3:1 ratio between day and night export.	To model the carbon requirements of the rhizoids, sucrose transport out of the model was added, and assumed to occur in a 3:1 ratio between day and night (ratio of day:night sugar export to the phloem in *Arabidopsis (312)*). Flux to the rhizoids was constrained such that it carried 4.3 % of the carbon assimilated by the model, based on data from *Polytrichum commune (152)*.
Diel transfer of starch	Constrained to 0.00127 (effective starch flux to biomass when all conditions are equal)	We constrained diel transfer of starch to a value based on the contents of starch in *Marchantia* biomass (*147*), based on evidence that *Marchantia* doesn't accumulate a large amount of starch (*148*).

Table S2.2: Primers used in this study.

Name	Sequence	Description
Sequencing		
AB_seq1	cggcaggatgcttaacataag	Sequencing primer for condensate-forming constructs, binding in mCherry
AB_seq2	tcttgctcgcctctatcatgc	Sequencing primer for condensate-forming constructs, binding in IntF2A
AB_seq3	cagcagcagtaacaaactcg	Sequencing primer for condensate-forming constructs, binding in mClover3
AB_seq4	gttcctcacctttgctcac	
M13F	gtaaaacgacggccagt	In-house sequencing primer M13F, binding in cloning vector backbone
M13R	caggaaacagctatgac	In-house sequencing primer M13R, binding in cloning vector backbone
AS_seq_09F_alt	ggaagttcatttcatttggaga	Sequencing primer for pK7WG2 vector binding in p35S. Courtesy of András Sándor.
AS_seq_011R	aggtcactggattttggttttag	Sequencing primer for pK7WG2 vector binding in T35S. Courtesy of András Sándor.
ALB_seq_FLuc	cagcctaatgactagtatggaa	Sequencing primer for PULSE binding in FLuc
ALB_seq_PhyB	gagaacaagctcaatcgtcg	Sequencing primer for PULSE binding in PhyB
ALB_seq_RLuc	tgtcgcgaatggcttcgaaagt	Sequencing primer for PULSE binding in RLuc
ALB_seq_4Frag3_HygR	attcggcgttaattcagtac	Sequencing primer for PULSE_HygR binding in HygR
ALB_seq_HygR_4Frag1	caaattgacgcttagacaac	Sequencing primer for PULSE binding upstream of left border
ALB_seq_4Frag1_4Frag2	tcttccagcggatagaatgg	Sequencing primer for PULSE binding in FLuc
ALB_seq_4Frag2_4Frag3	cgatttcgatctggacatg	Sequencing primer for PULSE binding in VP16
ALB_seq_PhyBdel_fwd	gtggttacattcagcctttcg	Sequencing primer for PULSE binding in PhyB
ALB_seq_PhyBde	ctgtaaaccgaaagcctgc	Sequencing primer for PULSE

l_rev		binding in PhyB
ALB_seq_PhbA_PhbB	tctggtgaccttgctgcatg	Sequencing primer for ConstPHB_P2AT2A and ConstPHB_InteinF2A binding in PhbA
ALB_seq_PhbBjoin	agtctgaggtcggagaggtg	Sequencing primer for ConstPHB_P2AT2A binding in PhbB
ALB_seq_PhbC_PhbB	ctggggtcactttgaaaggc	Sequencing primer for ConstPHB_P2AT2A and ConstPHB_InteinF2A binding in PhbC
ALB_seq_PhbCnonsense_fwd	gacggaggctcgacataag	Sequencing primer for ConstPHB_InteinF2A binding in InteinF2A
ALB_seq_PhbCnonsense_rev	gagtgaactctcaggttg	Sequencing primer for ConstPHB_InteinF2A binding in PhbC
ALB_seq_ACLY3KQ_Int_AttDT	ctgggatgatatctcctacg	Sequencing primer for ACLY3KQ_IntF2A_AttDT binding in ACLY3KQ
ALB_seq_IntF2A_AttDT	acctgttaaagcttgctggc	Sequencing primer for ACLY3KQ_IntF2A_AttDT and *Att*DT_IntF2A_AttDT binding in InteinF2A
ALB_seq_MidAttDT	cttgtcggtatgtggaagtc	Sequencing primer for ACLY3KQ_IntF2A_AttDT and *Att*DT_IntF2A_AttDT binding in *Att*DT
ALB_seq_pMpEF1a_ACLY	gaggttggctgtgtgctatg	Sequencing primer for ACLY3KQ_IntF2A_AttDT and *Att*DT_IntF2A_AttDT binding in pMpEF1α
ALB_seq_ACLY_1	tacatcgctgatttggatgc	Sequencing primer for ACLY3KQ_IntF2A_AttDT binding in ACLY3KQ
ALB_seq_ACLY_2	ccttcactggtgatcataag	Sequencing primer for ACLY3KQ_IntF2A_AttDT binding in ACLY3KQ
ALB_seq_ACLY_3	ggagttatcgtgcctgctc	Sequencing primer for ACLY3KQ_IntF2A_AttDT binding in ACLY3KQ
ALB_seq_M13Ralt	gttagctcactcattaggcac	Alternative M13R sequencing primer binding in backbone of p*Mp*GWB Gateway vectors
ALB_seq_Frag3I_Frag4	acctacatttacgggagcag	Sequencing primer for PULSE_HygR_Phb binding in

		PhbC
Gibson assembly		
HygR_fwd	aattcttgcggaggagctaactcacattaggc	Primers used for the assembly of PULSE_HygR from p*Mp*GE010_g1_BHLH51 and PULSE
HygR_rev	cgtcgatctagtttgcgtattggctagagc	
JPUB4frag_1_fwd	agccaatacgcaaactagatcgacgctggatc	
JPUB4frag_1_rev	gcgatctgacggttcactaaacgagctctgc	
JPUB4frag_2_fwd	ctcgtttagtgaaccgtcagatcgcctggag	
JPUB4frag_2_rev	atcggtaaacatctgctcaaactcgaagtcgg	
JPUB4frag_3_fwd	cgagtttgagcagatgtttaccgatgccctg	
JPUB4frag_3_rev	gtgagttagctcctccgcaagaattcaagc	
Constr1_fwd	aggaagtgccctcttgtcgtcggccaaataatg	Primers used for the assembly of ConstPHB_P2AT2A from genes synthesized by TWIST
Constr1_rev	gaccccagccgcggtcagccattccgtc	
Constr2_fwd	ggctgaccgcggctggggtcggatcgtt	
Constr2_rev	atcgcagacttatcatcactgagcctccacc	
Backbone_fwd	agtgatgataagtctgcgatggtggatg	
Backbone_rev	acgacaagagggcacttccttattccag	
ConstrInt1_fwd	ataaggaagtgccctcttgtcgtcggccaaataatg	Primers used for the assembly of ConstPHB_InteinF2A from genes synthesized by TWIST
Constr1Int_rev	ggcaagcctgcttaagatcctcctcggataataagtttc	
Constr3Int_fwd	gaggaggatcttaagcaggcttgcctttctttc	
Constr3Int_rev	gtggtggtgatggtgatgatggtggtgatgc	
Constr2Int_fwd	ccaccatcatcaccatcaccaccacatggcc	
Constr2Int_rev	accatcgcagacttatcatcactgagcctccac	
BackboneInt_fwd	tcagtgatgataagtctgcgatggtggatg	
BackboneInt_rev	cgacgacaagagggcacttccttattccag	
ACLY3KQ_IntF2A_fwd	atgtgtcatgggtcatagctgtttcctgtg	Primers used for the assembly of ACLY3KQ_IntF2A_*Att*DT from genes synthesized by TWIST
ACLY3KQ_IntF2A_rev	ttaggatctcggtaccgaaagacaggcag	
IntF2A_AttDT_fwd	tttcggtaccgagatcctaacagtcgaatatg	

IntF2A_AttDT_rev	agctatgacccatgacacatgcatccac	
AtALB_fwd	agcaggctatgttttgaaatggcgaggaag	Primers used for the assembly of *At*ALB_IntF2A_AttDT from pDONR21_*At*ALB and ACLY3KQ_IntF2A_AttDT
AtALB_rev	tccgaaatgagcttctgctccttggtgtacaacacatc	
BB_AttDT_fwd	gagcagaagctcatttcg	
BB_AttDT_rev	atttcaaaacatagcctgctttttgtac	
PULSE_HygR_frag1_fwd	aattcgtaaggaggttactagatcgacgctgag	Primers used for the assembly of PULSE_HygR_Phb from PULSE_HygR and ConstPHB_InteinF2A
PULSE_HygR_frag1_rev	attgtggtgtaaacaaattgacgcttagacaac	
PULSE_HygR_frag2_fwd	agcgtcaatttgtttacaccacaatatatcctgcc	
PULSE_HygR_frag2_rev	atgcatagcctgactagtcattaggctggatc	
InteinF2A_allPHB_fwd	cctaatgactagtcaggctatgcatcatcacc	
InteinF2A_allPHB_rev	gaagccctgcaggttaagcgtagtccggaac	
PULSE_HygR_frag4_fwd	gactacgcttaacctgcagggcttctctag	
PULSE_HygR_frag4_rev	cgatctagtaacctccttacgaattcccatgg	
GA1F	cgaacaatgaacccagctttcttgtacaaag	Primers used for the assembly of GA2.1 (mClo3R and RGGF) and GA2.1noNLS (mClo3RnoNLS and RGGFnoNLS) from gene fragments synthesized by TWIST and pmClo3_FL (mClover3)
GA1R	ctttgctcacaccgtctccaccattgtc	
mClo3F	tggagacggtgtgagcaaaggtgaggaac	
mClo3R	gattgctttccactttacgctttttcttcg	
mClo3RnoNLS	gattgctttctttgtaaagttcgtccatacc	
RGGF	gcgtaaagtggaaagcaatcagagcaataac	
RGGFnoNLS	actttacaaagaaagcaatcagagcaataac	
RGGR	gcccgtccccaccattgtcccttcgatc	
mCheF	ggacaatggtggggacgggcatcaccac	
mCheR	aaagctgggttcattgttcgtgagaggccac	
GA3_1F	tgacgataaggtgagcaaaggtgaggaac	Primers used for the assembly of GA3 (GA3_1R and GA3_2F) and GA3noNLS (GA3_1noNLSR and GA3_2noNLSF) from GA2.1 and GA2.1noNLS
GA3_1R	gatgatgatgcactttacgctttttcttcgg	
GA3_1noNLSR	gatgatgatgtttgtaaagttcgtccatacc	
GA3_2F	gcgtaaagtgcatcatcatcatcatcactgc	
GA3_2noNLSF	actttacaaacatcatcatcatcatcact	

	gc	
GA3_2R	ctttgctcaccttatcgtcatcatctttataatcttc	
MdCMS_fwd	tcacgttagcatggccttctcagcagaaaatc	Primers used for the assembly of GA2.1noNLS_MdCMS (MdCMS_fwd and BB_rev) and GA2.1noNLS_MdCMSnoChl (MdCMSnoChl_fwd and BBnoChl_rev)
MdCMSnoChl_fwd	tcacgttagccccactcaccccgaatac	
MdCMS_rev	tatgagatgtgatgtccaacttgatgttgac	
BB_fwd	gttggacatcacatctcataacgaggattac	
BB_rev	agaaggccatgctaacgtgatgatgatg	
BBnoChl_rev	ggtgagtggggctaacgtgatgatgatg	
RGG_mClover3_fwd	tgaggaacttttcactggggttgtgcctatc	Primers used for the assembly of SYNZIP1_RGG_mClo3 from GA2.1noNLS
RGG_mClover3_rev	gctgggttcaatgatgatgatggtggtgatgc	
mClover3_RGG_BB_fwd	tcatcatcattgaacccagctttcttgtac	
mClover3_RGG_BB_rev	ccccagtgaaaagttcctcacctttgctc	
mClo_fwd	tgacgataaggtgagcaaaggtgaggaac	Primers used for assembly of SYNZIP1_mClo3 from SYNZIP1_RGG_mClo3
mClo_rev	ggtggtgatgtttgtaaagttcgtccatacc	
ZIP1BB_fwd	actttacaaacatcaccaccatcatcatc	
ZIP1BB_rev	ctttgctcaccttatcgtcatcatctttataatc	
SYNZIP2_PhbB2546_fwd	taggtggaggctcagtgatgataagtctg	Primers used for the assembly of SYNZIP2_PhbABC from gene fragments synthesized by TWIST
SYNZIP2_PhbB2546_rev	cctagtaatacctgcattattaatcaagac	
SYNZIP2_PhbB2547_fwd	ataatgcaggtattactagggacgttgtcttc	
SYNZIP2_PhbB2547_rev	catcactgagcctccacctagcctcaaataatg	
BB1_fwd	ctcgacataagcaaaagatagtggcacc	Primers used for the assembly of SYNZIP2_PhbAmCherry_BC (BB1PhbA_rev, mCherryLinkerPhbA_fwd, BB2PhbA_fwd), SYNZIP2_PhbABmCherry_C (BB1PhbB_rev, mCherryLinkerPhbB_fwd, BB2PhbB_fwd), and SYNZIP2_PhbABCmCherry (BB1PhbC_rev,
BB1PhbA_rev	ccgtagctccgtgatgatgttgttcgtg	
BB1PhbB_rev	ccgtagctccatggtggtgttgttcgtg	
BB1PhbC_rev	ccgtagctccagccttggccttaacgtatc	
mCherryLinkerPhbA_fwd	acatcatcacggagctacggccagtgct	
mCherryLinkerPhbB_fwd	acaccaccatggagctacggccagtgct	
mCherryLinkerPhbC_fwd	ggccaaggctggagctacggccagtgct	

mCherryLinkerPhbA_rev	caacgtcggtatcgtagtacgacattgcgcc	mCherryLinkerPhbC_fwd, BB2PhbC_fwd) from pK7WG2_MdCMS_mC_SpC and SYNZIP2_PhbABC
mCherryLinkerPhbB_rev	ccttatagtcatcgtagtacgacattgcgcc	
mCherryLinkerPhbC_rev	cgtaaggataatcgtagtacgacattgcgcc	
BB2PhbA_fwd	gtactacgataccgacgttgtgattgtc	
BB2PhbB_fwd	gtactacgatgactataaggatgatgacgataaaatg	
BB2PhbC_rev	gtactacgattatccttacgatgttccgg	
BB2_rev	tatcttttgcttatgtcgagcctccgtc	

Table S2.3: Plasmids used in this study. PhbA: β-ketothiolase, PhbB: acetoacetyl-CoA reductase, PhbC: PHB synthase.

Name	Description	Source
pK7WG2	Gateway expression vector for overexpression in plants using the Cauliflower Mosaic Virus 35S promoter (p35S). Carries a kanamycin resistance gene.	Gift from Francesco Licausi
p*Mp*GWB103	Gateway expression vector for overexpression in *M. polymorpha* using the p*Mp*EF1α promoter. Carries a hygromycin resistance gene.	Gift from Takayuki Kohchi
p*Mp*GWB303	As p*Mp*GWB103 but carries a chlorsulfuron resistance gene.	Gift from Takayuki Kohchi
p*Mp*GE010_g1_BHLH51	Source of hygromycin resistance cassette used for cloning of PULSE_HygR	Gift from Francesco Licausi
PULSE	PULSE optogenetic system, carries kanamycin resistance gene.	Gift from Matias Zurbriggen
PULSE_HygR	As PULSE but with a hygromycin resistance gene.	This study
PULSE_HygR_Phb	As PULSE_HygR but firefly luciferase coding sequence replaced with the cassette coding for PhbA, PhbB, and PhbC separated by the InteinF2A viral self-cleaving peptide.	This study
ConstPHB_P2AT2A	Cloning vector containing the cassette coding for PhbA, PhbB, and PhbC separated by the P2A and T2A viral self-cleaving peptides	This study
ConstPHB_InteinF2A	Cloning vector containing the cassette coding for PhbA, PhbB, and PhbC separated by the InteinF2A viral self-cleaving peptide	This study
GWB103_ConstPHB_P2AT2A	As ConstPHB_P2AT2A but in the p*Mp*GWB103 expression vector	This study
GWB103_ConstPHB	As ConstPHB_InteinF2A but in the p*Mp*GWB103	This study

_InteinF2A	expression vector	
ACLY3KQ_IntF2A_AttDT	Cloning vector containing the cassette coding for the human ACLY with three K>Q mutations and *At*tDT separated by the InteinF2A viral self-cleaving peptide	This study
pK7_ACLY3KQ_IntF2A_AttDT	As ACLY3KQ_IntF2A_AttDT but in the pK7WG2 expression vector	This study
GWB303_ACLY3KQ_IntF2A_AttDT	As ACLY3KQ_IntF2A_AttDT but in the p*Mp*GWB303 expression vector	This study
pDONR21_*At*ALB	Source of *At*ALB coding sequence	Gift from Yves Poirier
***At*ALB_IntF2A_AttDT**	Cloning vector containing the cassette coding for the chimeric ACLY *At*ALB and *At*tDT separated by the InteinF2A viral self-cleaving peptide	This study
pK7_*At*ALB_IntF2A_AttDT	As *At*ALB _IntF2A_AttDT but in the pK7WG2 expression vector	This study
GWB303_*At*ALB_IntF2A_AttDT	As *At*ALB _IntF2A_AttDT but in the p*Mp*GWB303 expression vector	This study
GA2.1	Cloning vector containing the cassette coding for the RGG-based scaffold fused to mClover3 and SYNZIP1 for the formation of biomolecular condensates, and the client mCherry-SYNZIP2 separated by the InteinF2A viral self-cleaving peptide. Nuclear localized.	This study
GA3	As GA2.1 but with RGG domains removed.	This study
GA2.1noNLS	As GA2.1 but with nuclear localization signal (NLS) removed.	This study
GA3noNLS	As GA3 but with NLS removed.	This study
pK7_GA2.1	As GA2.1 but in the pK7WG2 expression vector.	This study
pK7_GA3	As GA3 but in the pK7WG2 expression vector.	This study
pK7_GA2.1noNLS	As GA2.1noNLS but in the pK7WG2 expression vector.	This study
pK7_GA3noNLS	As GA3noNLS but in the pK7WG2 expression vector.	This study
GWB103_GA2.1	As GA2.1 but in the p*Mp*GWB103 expression vector.	This study
GWB103_GA2.1noNLS	As GA2.1noNLS but in the p*Mp*GWB103 expression vector.	This study
pmClo3_FL	Source of mClover3 coding sequence	Gift from Francesco Licausi
pK7WG2_MdCMS_mC_SpC	Source of *Md*CMS coding sequence	Gift from András Sándor
GA2.1noNLS_MdCM	As GA2.1noNLS but with *Md*CMS added	This study

S	upstream and in frame with mCherry.	
GA2.1noNLS_MDCMSnoChl	As GA2.1noNLS_MdCMS but with putative chloroplast localization signal removed from *Md*CMS	This study
pK7_GA2.1noNLS_MdCMS	As GA2.1noNLS_MdCMS but in pK7WG2 expression vector	This study
pK7_GA2.1noNLS_MDCMSnoChl	As GA2.1noNLS_MdCMSnoChl but in pK7WG2 expression vector	This study
SYNZIP1_RGG_mClo3	Cloning vector containing the cassette coding for the RGG-based scaffold fused to mClover3 and SYNZIP1 for the formation of biomolecular condensates.	This study
SYNZIP2_PhbABC	Cloning vector containing the cassette coding for PhbA-SYNZIP2, PHB-SYNZIP2, and PhbC-SYNZIP2 fusion proteins separated by the InteinF2A viral self-cleaving peptide.	This study
SYNZIP2_PhbAmCherry_BC	As SYNZIP2_PhbABC but with mCherry fused to PhbA.	This study
SYNZIP2_PhbABmCherry_C	As SYNZIP2_PhbABC but with mCherry fused to PhbB.	This study
SYNZIP2_PhbABCmCherry	As SYNZIP2_PhbABC but with mCherry fused to PhbC.	This study
pK7_SYNZIP1_RGG_mClo3	As SYNZIP1_RGG_mClo3 but in the pK7WG2 expression vector	This study
pK7_SYNZIP2_PhbABC	As SYNZIP2_PhbABC but in the pK7WG2 expression vector.	This study
pK7_SYNZIP2_PhbAmCherry_BC	As SYNZIP2_PhbAmCherry_BC but in the pK7WG2 expression vector.	This study
pK7_SYNZIP2_PhbABmCherry_C	As SYNZIP2_PhbABmCherry_C but in the pK7WG2 expression vector.	This study
pK7_SYNZIP2_PhbABCmCherry	As SYNZIP2_PhbABCmCherry but in the pK7WG2 expression vector.	This study

Table S3.1: Convergence diagnostics of 5,000 samples of the baseline and cytosolic PHB-producing models using the OPTGP algorithm. Four sample sets (chains) were computed using the indicated models and thinning factors. N_{max} is an estimate of the total number of samples required for a set of samples to achieve convergence. The dependence factor (I) is indicative of autocorrelation (the degree of dependence between consecutive samples). Chains with I > 5 are considered to show high dependence. Both N_{max} and I are part of the Raftery & Lewis diagnostic. The Gweke diagnostic (z) tests whether the mean of the first samples of the chain is significantly different from the mean of the last samples. % I > 5 and % z > 1.28 indicate the percentage of reactions which do not pass the Raftery & Lewis and Gweke convergence diagnostics, respectively. Thresholds were chosen based on (*127*).

Model	Thinning constant	% I > 5 (Raftery & Lewis)	N_{max} (Raftery & Lewis)	% z > 1.28 (Gweke)
Baseline	1,000	27	698034	47
	10,000	16	322868	30
	100,000	5	69840	24
	1,000,000	0	27858	20
Cytosolic PHB	1,000	23	241615	26
	10,000	17	164820	42
	100,000	4	75257	35
	1,000,000	0	58905	14

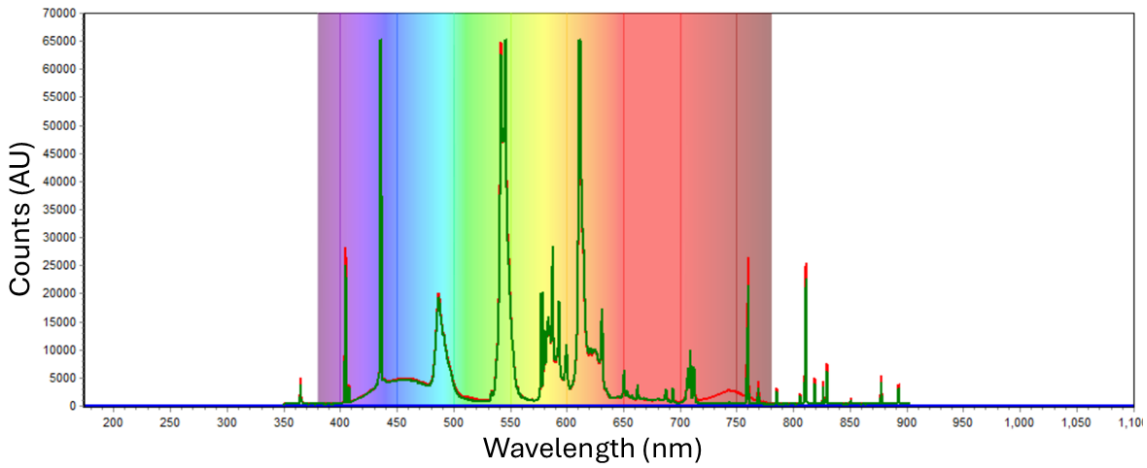


Figure S2.1: Superimposed light spectra of conditions used for growth of *M. polymorpha* in white light without (green) and with (red) supplementary far-red light.

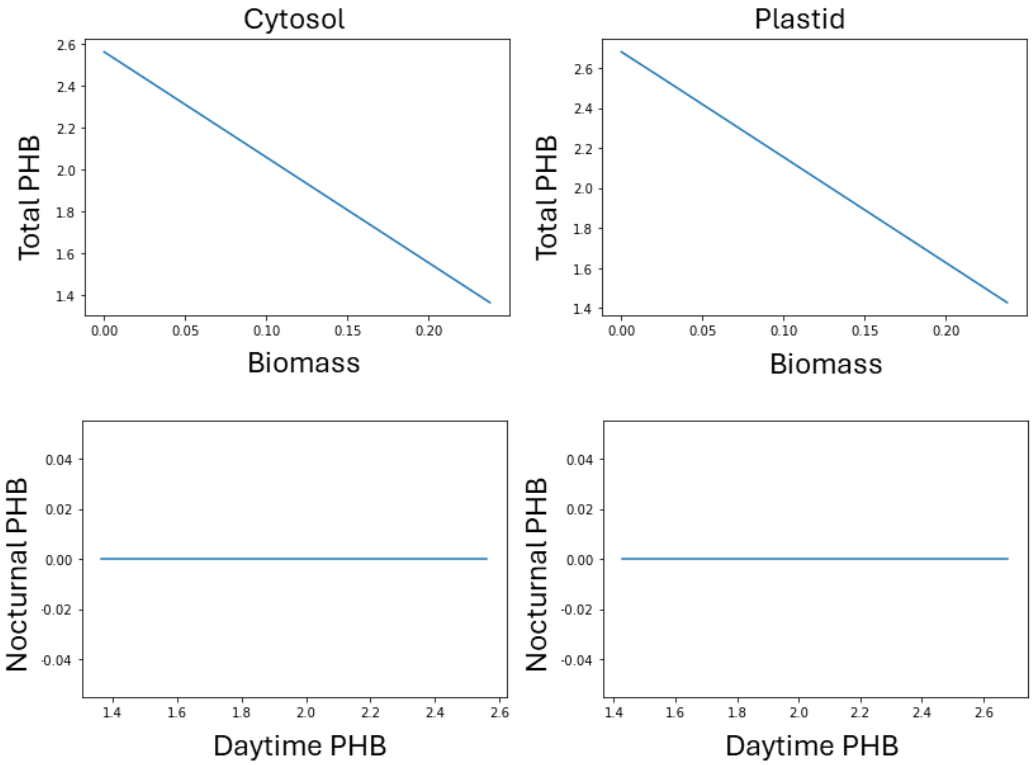


Figure S3.1: Pareto analysis of total PHB production versus biomass accumulation (top), and relative contributions to total PHB by night and daytime PHB production (bottom). All fluxes are in units of μmol m^{-2} s^{-1}.

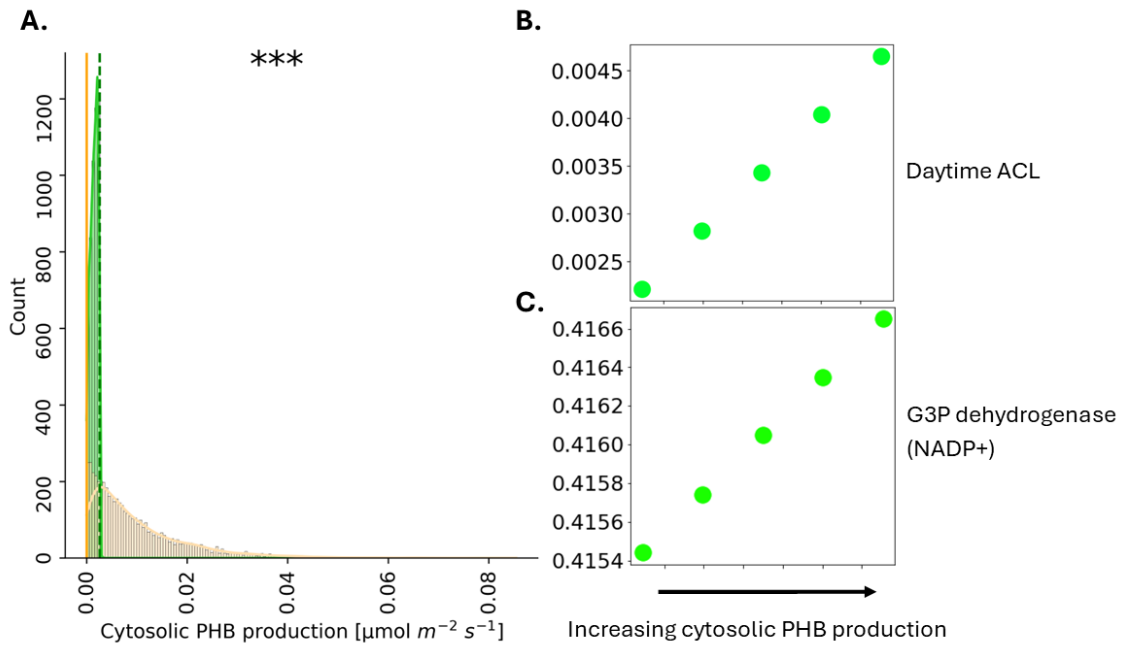


Figure S3.2: Flux sampling and correlation analysis of the cytosolic PHB accumulation reaction. (A) The flux distributions of the cytosolic PHB accumulation reaction in the baseline (orange) and cytosolic PHB-producing (green) models. Result of Kruskal-Wallis analysis of difference between distributions shown at the top of the figure. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05. Green dotted and orange lines represent the value carried by the reaction following pFBA in the cytosolic PHB and baseline models, respectively. (B-C) Results of correlation analysis between cytosolic PHB accumulation and daytime ACL (B), and daytime GAPN (C). All fluxes are in units of µmol m^{-2} s^{-1}.

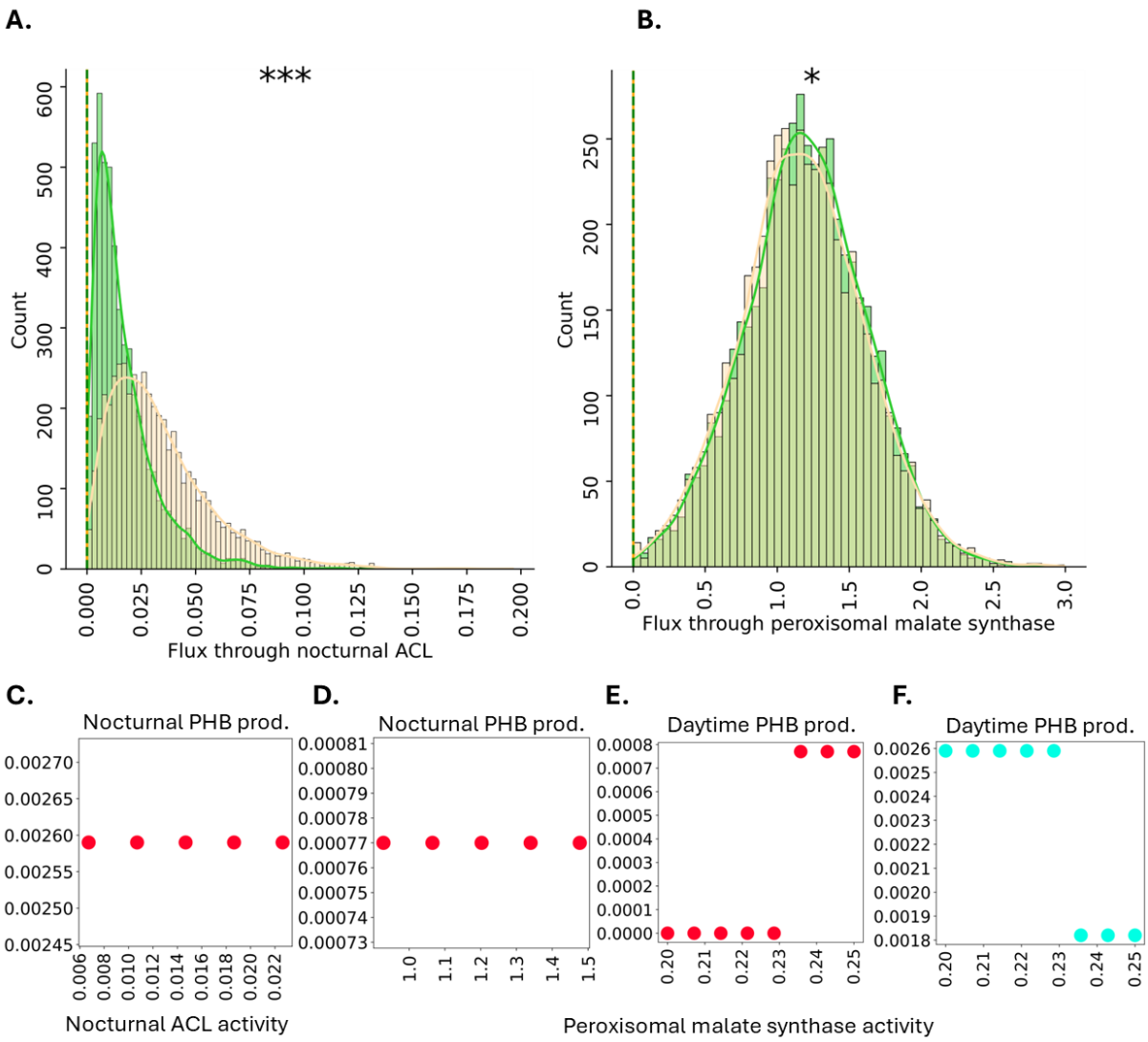


Figure S3.3: Flux sampling and correlation analysis of the nocturnal ACL and peroxisomal malate synthase reactions. The flux distributions of the nocturnal ACL and peroxisomal malate synthase reactions (A and B, respectively) in the baseline (orange) and cytosolic PHB-producing (green) models. Result of Kruskal-Wallis analysis of difference between distributions shown at the top of the figure. '***' represents a *P*-value < 0.001, '**' a *P*-value < 0.01, '*' a *P*-value < 0.05, or 'n.s.' a *P*-value > 0.05. Green dotted and orange lines represent the value carried by the reaction following pFBA in the cytosolic PHB and baseline models, respectively. (C) Results of correlation analysis between nocturnal ACL and nocturnal PHB production. (D-F) Results of correlation analysis between peroxisomal malate synthase and (D) nocturnal PHB production, (E) daytime PHB production around the critical value, and (F) nocturnal PHB production around the critical value. All fluxes are in units of μmol m^{-2} s^{-1}.

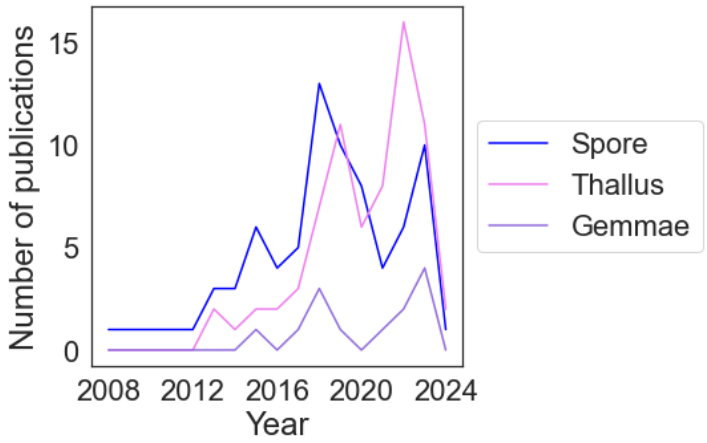


Figure S4.1: The number of publications per year making use of spore, thallus, or gemmae-based *M. polymorpha* transformation methods. Data obtained from Clarivate™ (Web of Science™). © Clarivate 2024. All rights reserved.

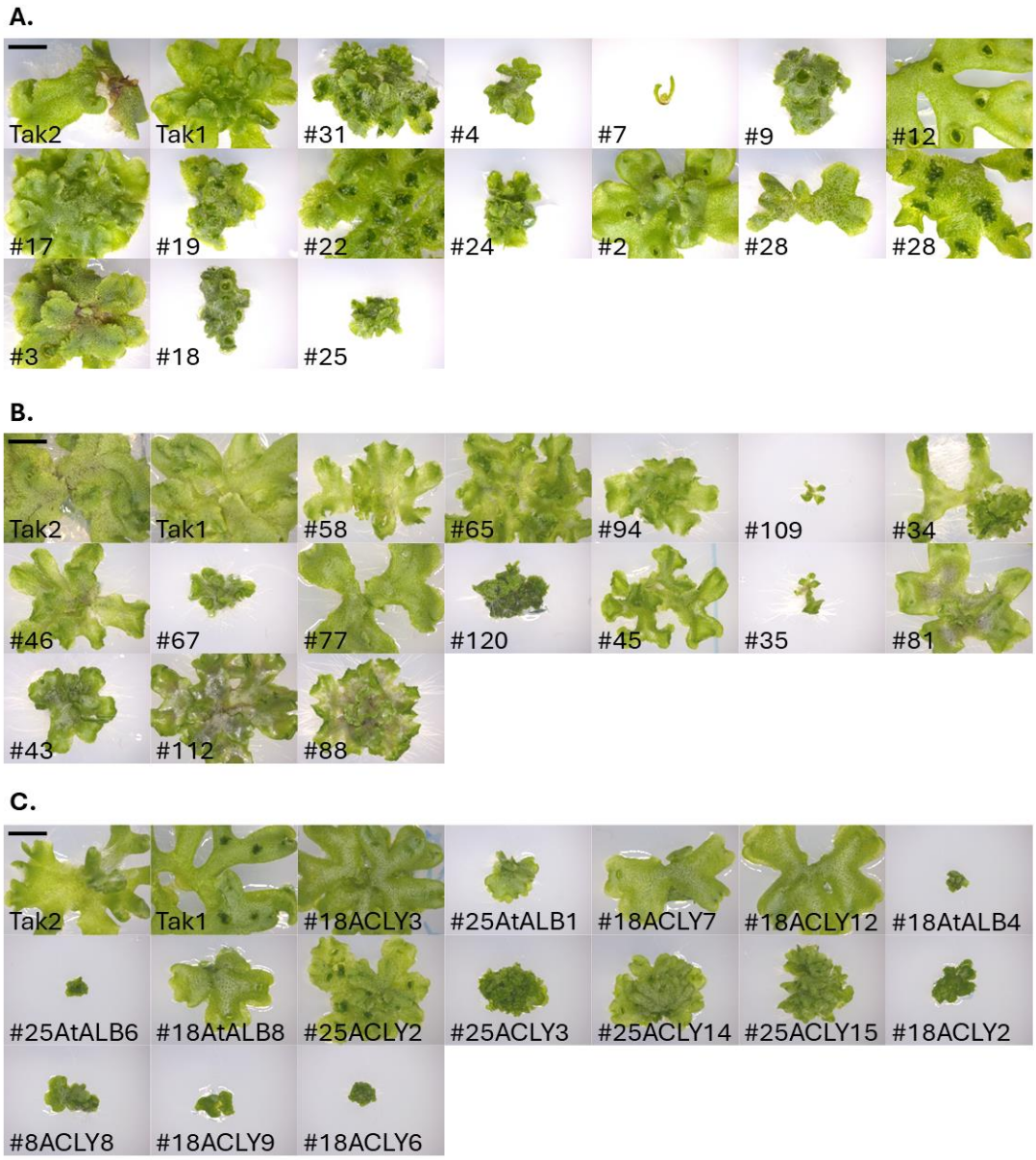


Figure S5.1: An illustration of the different phenotypes obtained in screens of *M. polymorpha* expressing (A) constitutive PHB, (B) inducible PHB, and (C) ancillary intervention constructs. Representative scale bar of 5 mm is included for each set.

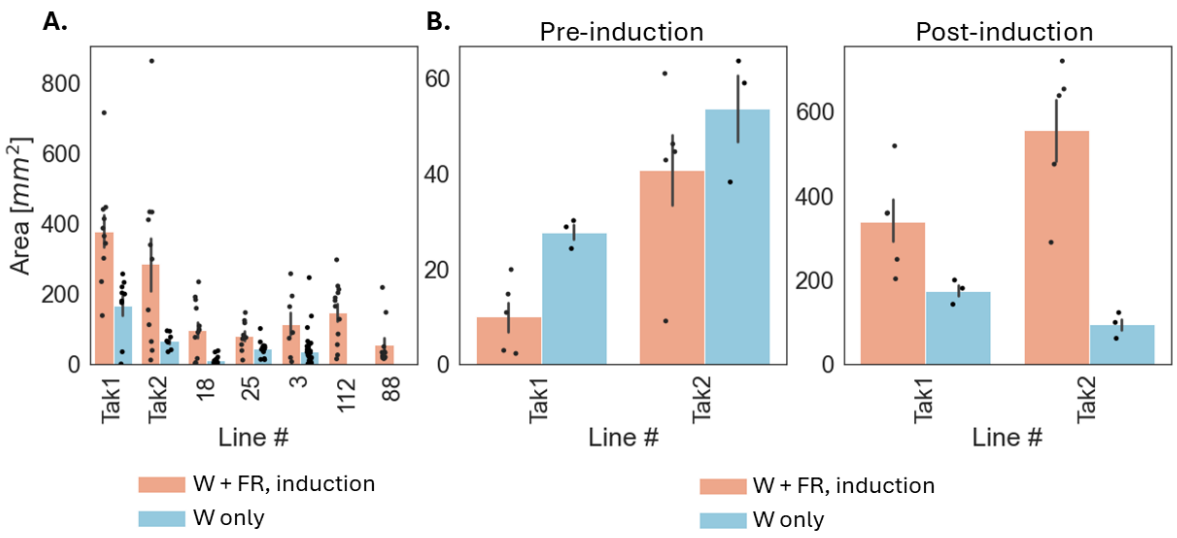


Figure S5.2: Impact on plant growth of induction conditions used in this study. (A) Final area of WT, constitutive PHB (lines #18, 25, 3) and inducible PHB (lines #112, 88) *M. polymorpha* plants after growth for 31 d. Plants were grown either in white light (W), or in white light supplemented with far-red light (W + FR) and induction according to scheme shown in Figure 5.2F. n = 7-31. Error bars represent SEM. (B) Comparison of size of WT Tak1 and Tak2 plants when (left) grown for 16 d in either W or W + FR conditions, and (right) after another 15 d during which some plants were subjected to three three-day bursts of red light induction. n = 3-5. Error bars represent SEM.

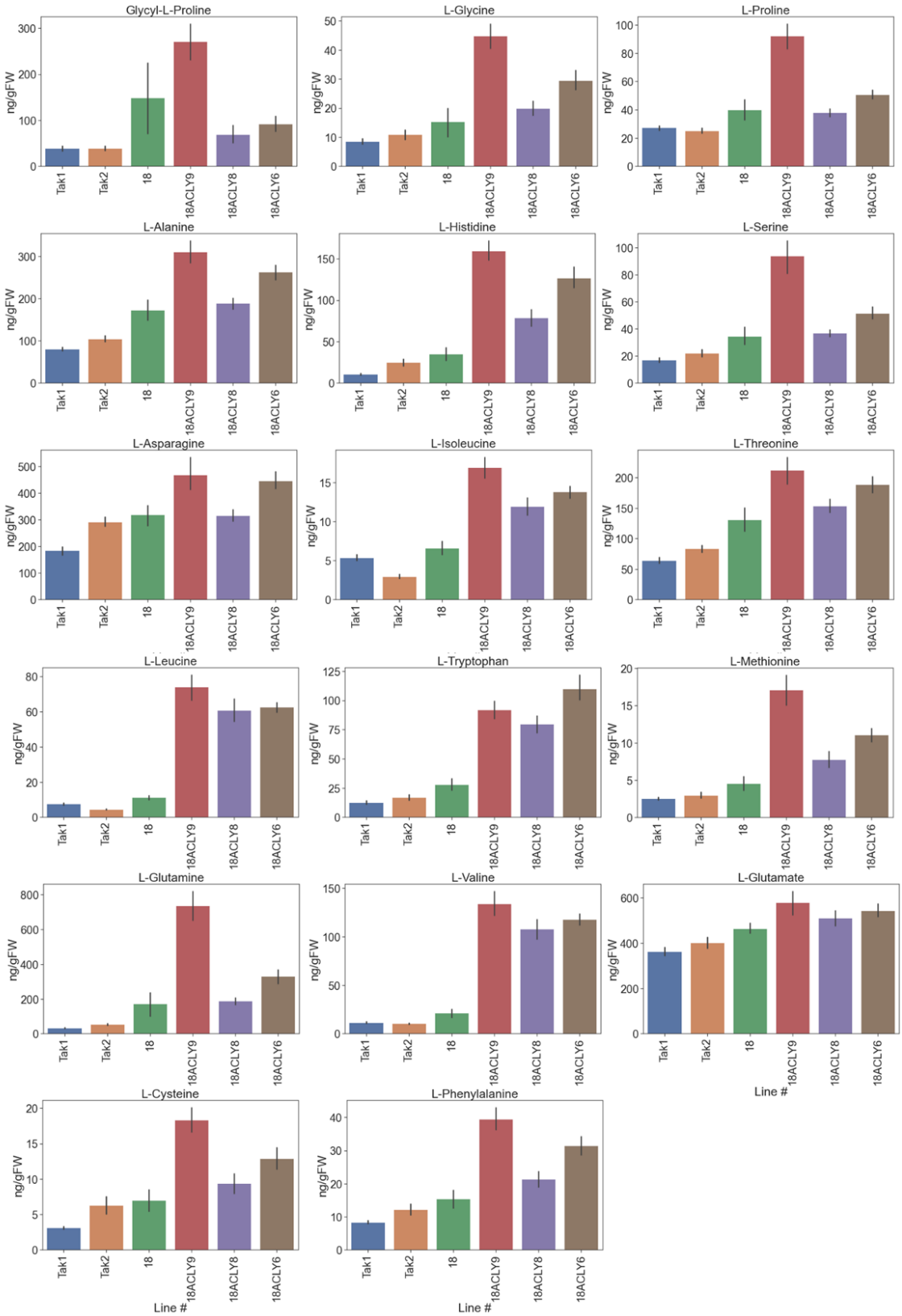


Figure S5.3: Results of the metabolite screen. Error bars indicate SEM.

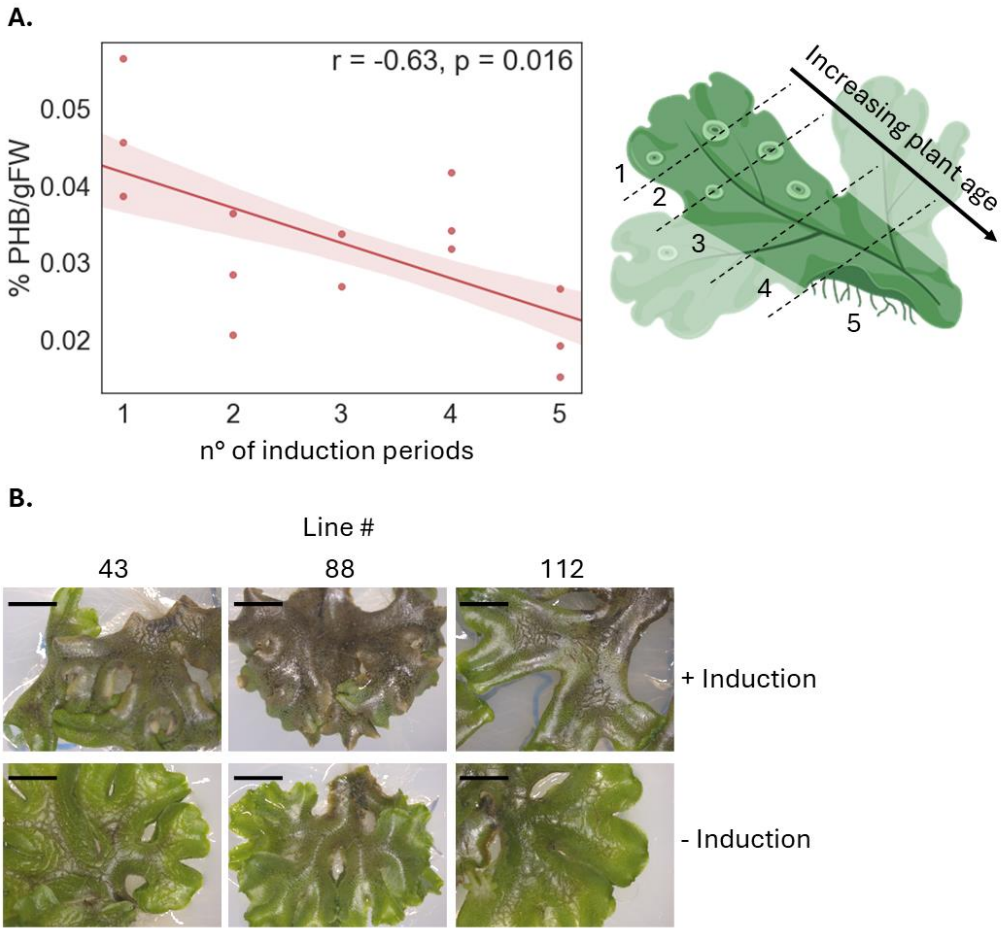


Figure S5.4: The effect of PHB accumulation on *M. polymorpha* thallus cuttings. (A) Sections of tissue having been exposed to increasing numbers of induction cycles according to the scheme shown in Figure 5.3C were collected and PHB extracted. This analysis was only conducted for inducible PHB line #112. Shaded area represents SEM. Pearsons correlation coefficient (r) and *P*-value (p) shown in top-right corner. n = 3. (B) Phenotypes of inducible PHB lines grown for 31 d with (+) or without (-) induction. Scale bar is 5 mm.

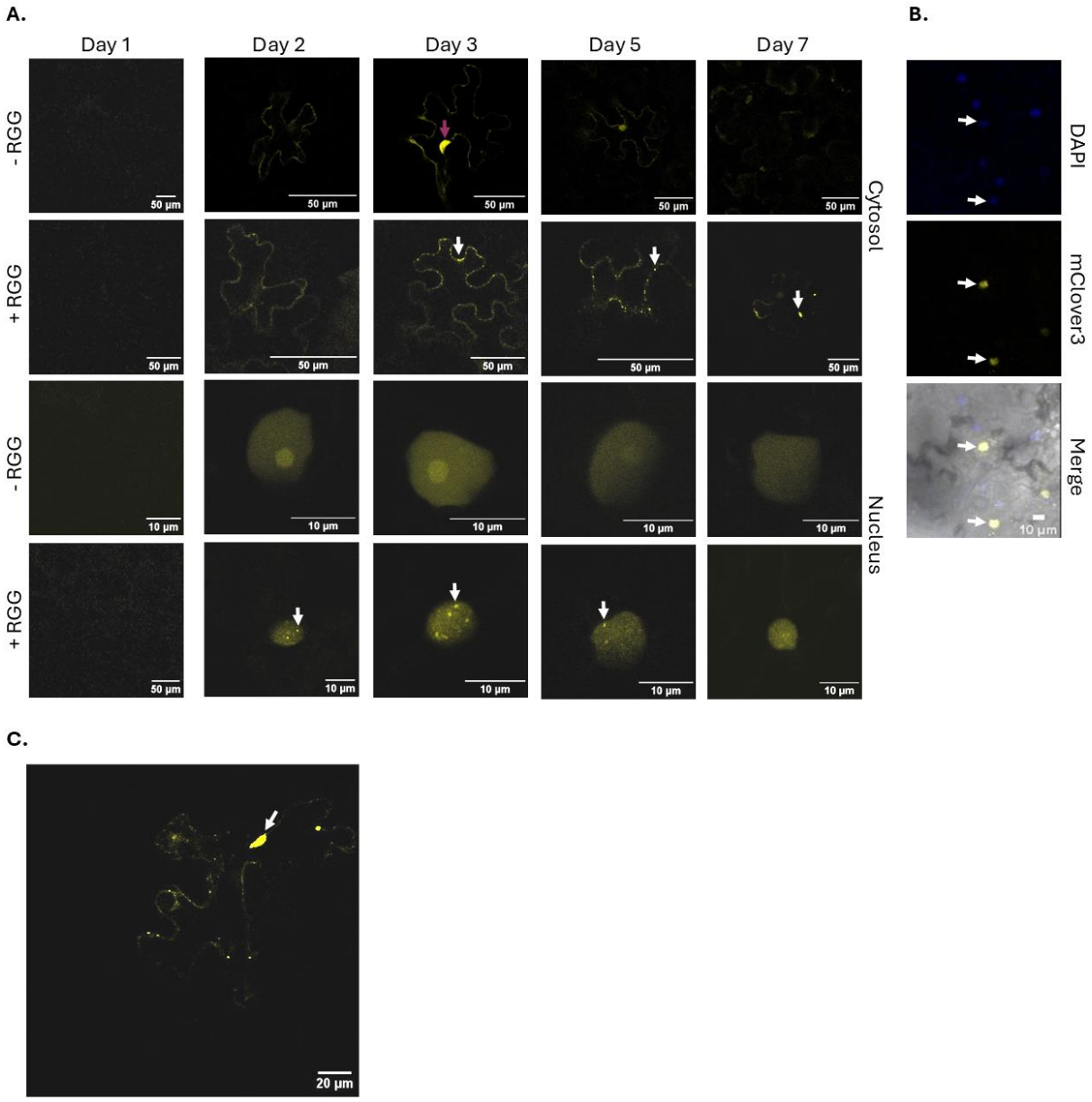
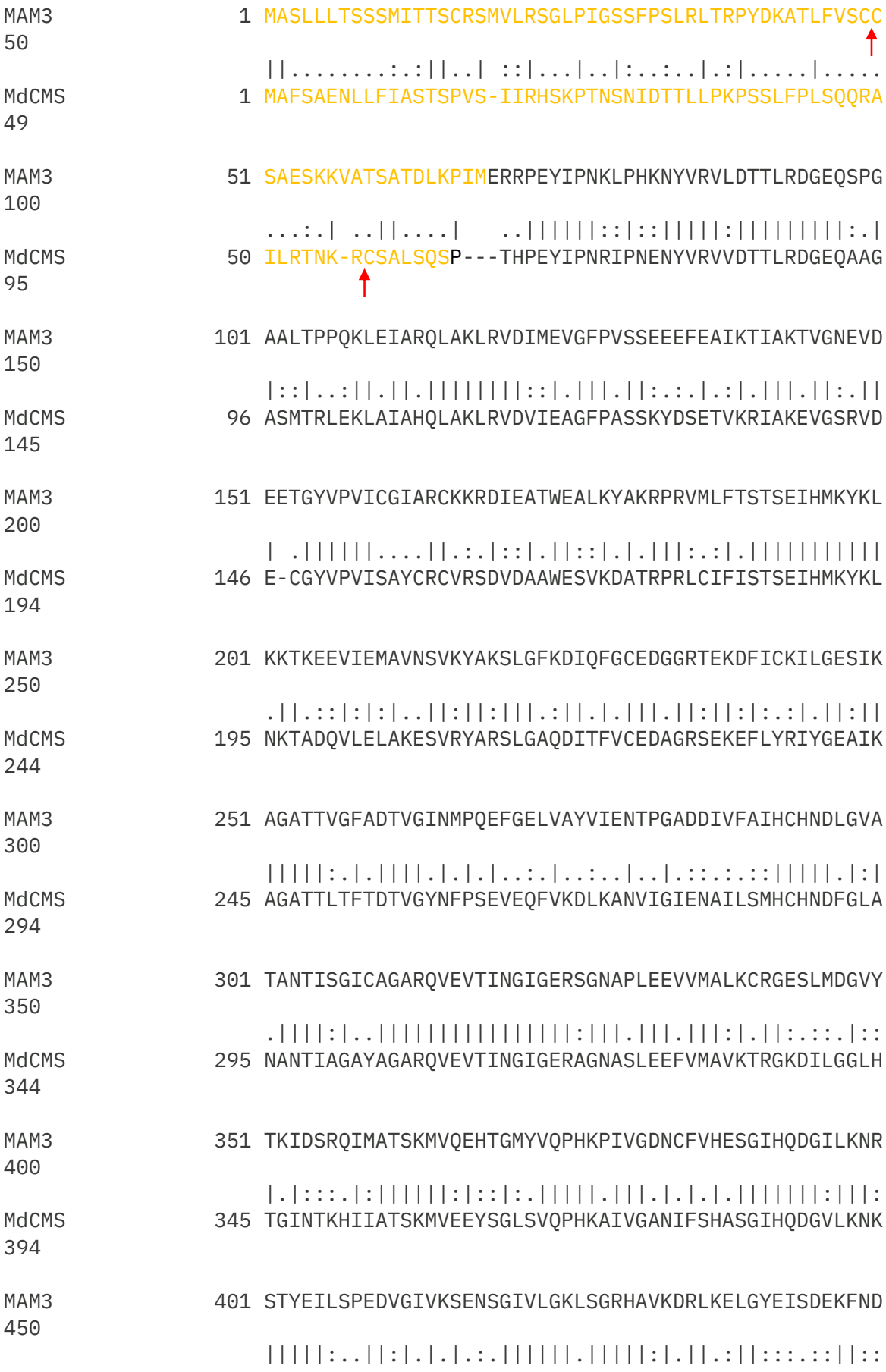


Figure S6.1: Further characterisation of biomolecular condensates *in planta*. (A) Timecourse of expression over 7 dpi in nuclei and cytosol of tissue infiltrated with the RGG-containing or no-RGG-containing version of the scaffold. mClover3 signal is shown. White arrows indicate condensates. Purple arrow indicates a nucleus; the presence of signal in the nuclei of cells overexpressing a FP is common. (B) DAPI-staining co-localises with mClover3 signal when an NLS is included, indicating that mClover3 gets targeted to the nucleus. (C) Very large biomolecular condensate indicated using a white arrow in a cell 7 dpi of the cytosolic biomolecular condensate-forming construct. mClover3 signal is shown.


MAM3 1 MASLLLTSSSMITTSCRSMVLRSGLPIGSSFPSLRLTRPYDKATLFVSCC 50
MdCMS 1 MAFSAENLLFIASTSPVS-IIRHSKPTNSNIDTTLLPKPSSLFPLSQQRA 49
MAM3 51 SAESKKVATSATDLKPIMERRPEYIPNKLPHKNYVRVLDTTLRDGEQSPG 100
MdCMS 50 ILRTNK-RCSALSQSP---THPEYIPNRIPNENYVRVVDTTLRDGEQAAG 95
MAM3 101 AALTPPQKLEIARQLAKLRVDIMEVGFPVSSEEEFEAIKTIAKTVGNEVD 150
MdCMS 96 ASMTRLEKLAIAHQLAKLRVDVIEAGFPASSKYDSETVKRIAKEVGSRVD 145
MAM3 151 EETGYVPVICGIARCKKRDIEATWEALKYAKRPRVMLFTSTSEIHMKYKL 200
MdCMS 146 E-CGYVPVISAYCRCVRSDVDAAWESVKDATRPRLCIFISTSEIHMKYKL 194
MAM3 201 KKTKEEVIEMAVNSVKYAKSLGFKDIQFGCEDGGRTEKDFICKILGESIK 250
MdCMS 195 NKTADQVLELAKESVRYARSLGAQDITFVCEDAGRSEKEFLYRIYGEAIK 244
MAM3 251 AGATTVGFADTVGINMPQEFGELVAYVIENTPGADDIVFAIHCHNDLGVA 300
MdCMS 245 AGATTLTFTDTVGYNFPSEVEQFVKDLKANVIGIENAILSMHCHNDFGLA 294
MAM3 301 TANTISGICAGARQVEVTINGIGERSGNAPLEEVVMALKCRGESLMDGVY 350
MdCMS 295 NANTIAGAYAGARQVEVTINGIGERAGNASLEEFVMAVKTRGKDILGGLH 344
MAM3 351 TKIDSRQIMATSKMVQEHTGMYVQPHKPIVGDNCFVHESGIHQDGILKNR 400
MdCMS 345 TGINTKHIIATSKMVEEYSGLSVQPHKAIVGANIFSHASGIHQDGVLKNK 394
MAM3 401 STYEILSPEDVGIVKSENSGIVLGKLSGRHAVKDRLKELGYEISDEKFND 450

```
MdCMS             395 STYEIILAEDIGYVHSNDDGIVLGKHSGRHALKSRLLQLGHDLDEKKFHE
444

MAM3              451 IFSRYRELTKDKKRITDADLKALVVNGAEISSEKLNSKGINDLMSSPQIS
500
                      :|..::.|.:.||.:|:.||::||...|
MdCMS             445 VFEHFKSLAETKKSLTNEDLESLVYQVA----------------------
472

MAM3              501 AVV    503
                        :
MdCMS             473 --I    473
```

Figure S6.2: Alignment between *Md*CMS and MAM3 from *Arabidopsis* used to determine whether *Md*CMS contains a chloroplast transit peptide. N-terminal unstructured region indicated in orange and predicted cleavage sites shown using a red arrow.