Linking the gene regulatory network with the functional physical structure of whole-genome engineered Arabidopsis mutants : an HR-MAS NMR-based metabolomics approach

Augustijn, D.

Linking the gene regulatory network with the functional physical structure of whole-genome engineered *Arabidopsis* mutants

An HR-MAS NMR-based metabolomics approach

Dieuwertje Augustijn

ISBN: 978-94-6332-407-6

Cover: Cyanotype photograpic image of *Arabidopsis thaliana*. Cyanotype is a photographic printing process where you add potassium ferricyanide and ferric ammonium citrate together which reacts with UV light to make a blue print.

This research was financially supported by the BioSolar Cells consortium and Foundation for Fundamental Research on Matter (FOM). The project was carried out in the research programme of BioSolar Cells (BSC core project grant FOM23) and was part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO).

Linking the gene regulatory network with the functional physical structure of whole-genome engineered *Arabidopsis* mutants

An HR-MAS NMR-based metabolomics approach

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Prof. Mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties te
verdedingen op dinsdag 4 december 2018
klokke 15.00 uur

door

Dieuwertje Augustijn
geboren te Amsterdam, Nederland
2 januari 1991

Promotiecommissie

Promotors:	Prof. dr. H.J.M. de Groot
	Prof. dr. A. Alia
Overige leden:	Prof. dr. H.S. Overkleeft
	Prof. dr. M. Ubbink
	Prof. dr. P.J.J. Hooykaas
	Prof. dr. N. van Dam German Centre for Integrative Biodiversity Research, Leipzig
	Dr. J. Harbinson Wageningen University and Research

Read first the best books. The important thing for you is not how much you know, but the quality of what you know

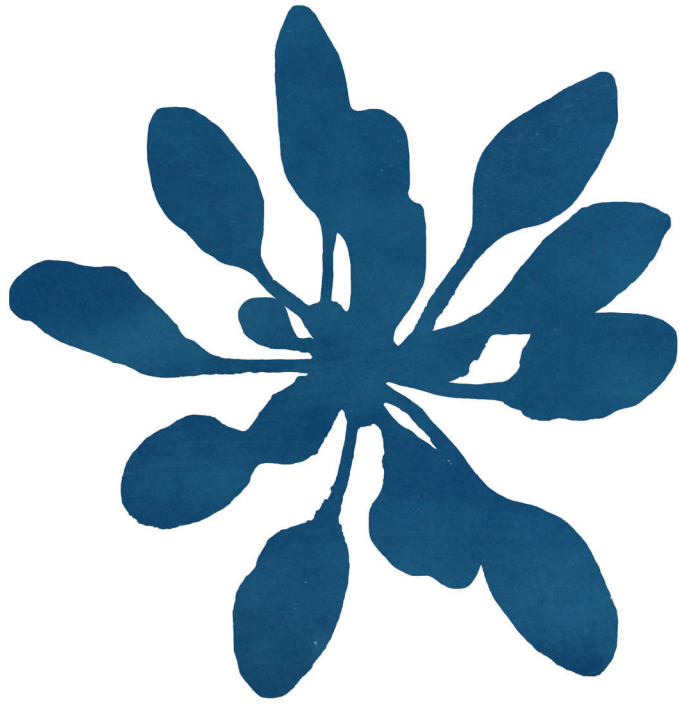
Erasmus

Table of Contents

Abbreviations

2DS	2 °C scenario
3F	Three zinc fingers
ATFs	Artificial transcription factors
BMRB	Biological magnetic resonance bank
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
CEA	Controlled-environmental agriculture
Chl	Chlorophyll
Col-0	Colombia O
COSY	Correlation spectroscopy
CPMAS	Cross polarization transfer
CPMG	Carr-Purcell-Meiboom- Gill
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/Cas9
CSI	Chemical shift imaging
d0	Increment delay
d1	Relaxation delay
d20	Spin echo delay
DDS	4-4-dimethyl-4-silapentane-1-sulfonic acid
DMF	N,N'-dimethylformamide
dpg	Days post germination
EC	Evening complex
ELF3	EARLY FLOWERING 3
ELF4	EARLY FLOWERING 4
FID	Free induction decay
FUM2	Fumerase 2
GABA	γ-aminobutyric acid
GC	Gas chromatography
HMBC	Heteronuclear multiple-bond correlation
HR-MAS	High-resolution magic angle spinning
HSQC	Heteronuclear single-quantum correlation
IEA	International Energy Agency
LC	Liquid chromatography

LCF1	Low Chlorophyll Fluorescence 1
LHY	LATE ELONGATED HYPOCOTYL
LUX	LUX ARRYTHMO
MAS	Magic angle spinning
mMDH1	Mitochondrial malate dehydrogenase 1
MORC2	Microrchidia family 2
MS	Mass spectrometry
MSH1	MutS HOMOLOGUE 1
MVA	Multivariate analysis
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
OPLS-DA	Orthogonal partial least squares discriminant analysis
PC	Principal component
PCA	Principal component analysis
PRR7	PSEUDO-RESPONSE REGULATORY7
RSA	Rosette surface area
SDGs	Sustainable development goals
SEM	Standard error
SLS	Slice localized spectroscopy
SUS	Shared and unique structures
TALENs	Transcription activator-like effector nucleases
TCA	Tricarboxylic acid
TMS	Tetramethylsilane
TOC1	TIME OF CAB EXPRESSION 1
TOCSY	Total correlation spectroscopy
TSP	3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid
ZF-ATF	Zinc finger artificial transcription factor
ZFs	Zinc-fingers



Introduction

The changing climate is a challenge for both current and future generations. To address these challenges, countries adopted the Paris agreement in 2015.[1] Two major goals of this agreement are that the global temperature increase stays well below 2 degrees Celsius above pre-industrial levels and to strive for an increase of not more than 1.5 degrees Celsius by the end of the century. It is also stated in the Paris agreement that the adaptions to accomplish the goals are not allowed to affect food production. This is also in agreement with the United Nations Sustainable Development Goals (SDGs) to end all forms of hunger and malnutrition in 2030 and to achieve food security.[2] Also, sustainable food production is promoted by the SDGs.

The challenge for the upcoming decades is to look for new resources, including biological resources, to meet both the demand in energy and food for the rising world population which is expected to reach 9.8 billion in 2050.[3] But in the meantime, it is also important to keep the 2 degree Celsius scenario (2DS) on track. In the 2DS, the goals are limiting the global temperature increase to 2 degree Celsius, but also reduce the CO_2 emission by 60% by 2050 in comparison to 2013.[4] One way of doing this is to make use of bioenergy, which is defined as the energy derived from the conversion of biomass.[5] Bioenergy can be used directly as fuel or processed into liquids or gases. The International Energy Agency (IEA) yearly examines the progress of renewable energy technologies to meet the 2DS targets. According to the recent report "Tracking clean energy progress 2017", the progress in bioenergy is not on track to reach the 2DS targets.[4] It is thus important to improve the yield of biomass resources, not only for food but also for bioenergy.

One way to produce more biomass is to enhance the productivity of agriculture. New scientific and technology tools can help in the challenge to produce sustainable agricultural products with increased yield and with a minimal environmental footprint. Such agricultural products are also called smart crops.[6,7]

Model plant organisms that can help in the development of smart crops are studied extensively by the research community.[8] The most used plant model is *Arabidopsis thaliana*. This is a small flowering rosette plant from the *Brassicaceae* family with a life cycle of 6 weeks. The plant has a small genome of 135 Mbp in total, which has been sequenced in 2000 as the first plant genome.[9,10] Transformations to obtain transgenic plants is an easy procedure.[11] *Arabidopsis* is widely used to understand molecular principles of plant development, cell biology, metabolism, physiology, genetics and epigenetics of plants. This is expanded in the last years to the field of systems biology.[9,10,12] The principal challenge in systems biology is to link the elements of the gene regulatory network with the functional physical structure and find novel molecular-based paths to plant development with enhanced yield. In this thesis, we will show a new unbiased method how this can be achieved.

1.1 Development of smart crop varieties

The development of smart crops can be accelerated by technologies to perform genome editing, in contrast to traditional breeding methods such as cross-breeding.[13] In particular, artificial transcription factors (ATFs) can be used to regulate the expression of target genes. Popular genome-editing tools using ATFs are clustered regularly interspaced short

palindromic repeats/Cas9 (CRISPR/Cas9), transcription activator-like effector nucleases (TALENs) and zinc-fingers (ZFs).[14-17]

In this study, *Arabidopsis* mutants are obtained using zinc finger artificial transcription factor (ZF-ATF)-mediated whole genome interrogation (Figure 1.1). This technology does not require a-priory knowledge or hypotheses and is therefore unbiased. New plants can be identified with rare mutations and phenotypes. ZF-ATFs consist of a DNA-binding domain from the zinc-finger proteins linked to an effector domain which is either an activator or a repressor domain.[14,16,18,19] The zinc protein is approximately 30 amino acids long, with a conserved ββα configuration and a backbone of conjugated cysteine (Cys) and histidine (His) residues and a zinc ion. The Cys_2-His_2 zinc finger domain is a very common DNA-binding motif in eukaryotes and binds to three base pairs in the major groove of the DNA.[14,16,18,19]

For the development of the mutants, an array of three zinc-fingers was used which recognizes the DNA sequence of 5'-GNN-3' (Figure 1.1). There are sixteen 5'-GNN-3'-binding ZFs available[20], which are leading to 4096 possible combinations of three zinc-fingers which will recognize 9 base pairs in the genome.[21] The effector domain used to develop the mutants for this study is the VP16 activator domain from the Herpes Simplex virus.[14,16,18,19] The artificial transcription factor library which originated by fusion of the 4096 three zinc-finger combinations with the VP16 domain is used to prepare transformed *Arabidopsis thaliana* plants using the floral dip method using *Agrobacterium*.[11,22]

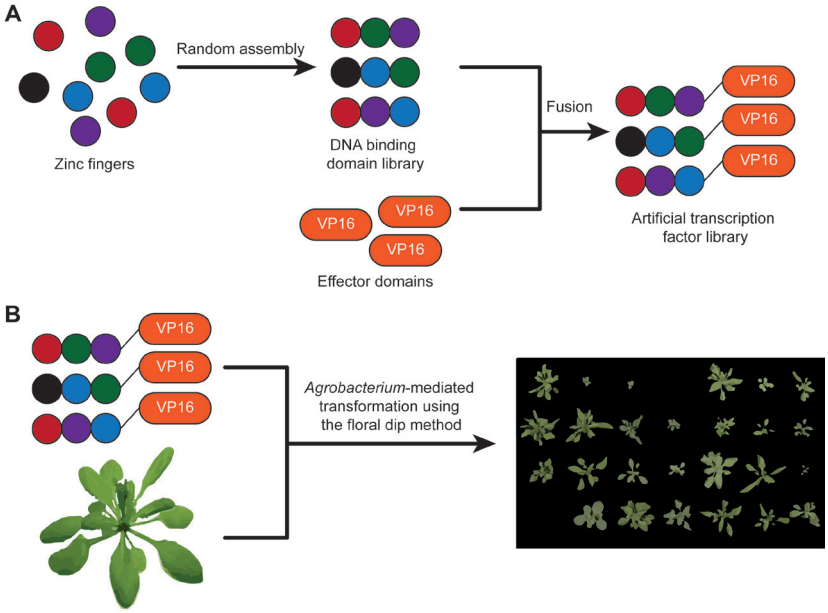


Figure 1.1: A) Development of the artificial transcription factor library by binding three zinc fingers together to prepare the DNA binding domain library and fuse this with the VP16 effector domain. B) Transformation of Arabidopsis thaliana plants with the artificial transcription factor library using the floral dip method to obtain Arabidopsis thaliana mutants with different phenotypes caused by the genome interrogation of the artificial transcription factors.

The first step in the floral dip method is to prepare a cell suspension of *Agrobacterium* strains including a binary vector containing the genes of the zinc-fingers. Then young flowering *Arabidopsis thaliana* plants are dipped into the *Agrobacterium* suspension to inoculate the plants with the *Agrobacterium*. The plants are grown in their normal growing conditions till the plants produce seeds. The seeds are collected and used to grow the first generation of transgenic plants (T_1).[11,22] These transgenic plants are screened for interesting novel phenotypes, such as growth enhancement[23], enhanced photochemistry[24] and increased tolerance to salinity.[25]

The purpose of this PhD thesis is to understand the novel mutant phenotypes from a systems biology perspective where omics technologies, like metabolomics and transcriptomics, are integrated with bio-imaging methods performed in a partner project.[26] The ultimate goal is to understand which pathways are leading to a specific phenotype so that it can be subsequently implanted to interesting crops for biomass production, such as plants of other *Brassicaceae* family that are used in bioenergy production[27] e.g. rapeseed (*Brassica napus*), Ethiopian mustard (*Brassica carinata*) and camelina (*Camelina sativa*).[28] And also to other plant families which are interesting for bioenergy production, such as maize (*Zea mays*), sugar beet (*Beta vulgaris*), sugarcane (*Saccharum officinarum*) and soybean (*Glycine max*).[29]

1.2 The role of metabolomics in systems biology

To understand the biological pathways underlying the novel mutant phenotypes, a systems biology approach can be used.[30-32] In systems biology, the information and interaction of the functional physical structure and the genetic information are integrated to provide a comprehensive model of the organism (Figure 1.2A). Different high-throughput technologies are used to study the genetic program of the various -omics fields: genomics, transcriptomics, proteomics and metabolomics (Figure 1.2B). This is complemented with information from

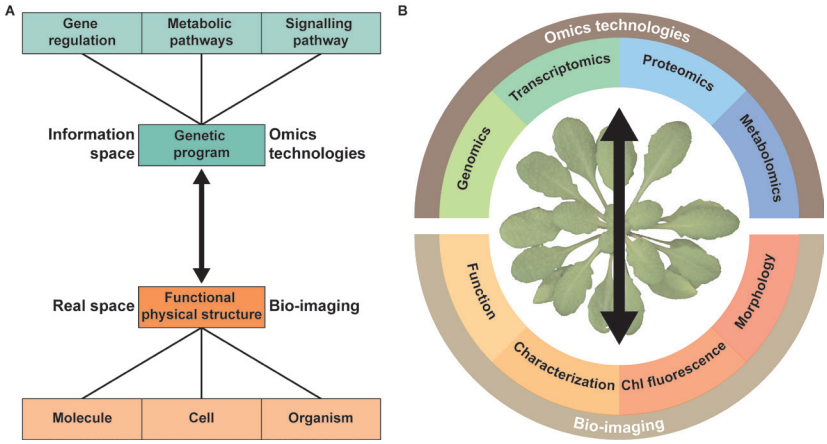


Figure 1.2: A) In systems biology, the information from the genetic program is integrated with information from functional physical structures to provide a comprehensive model of Arabidopsis thaliana. B) The genetic program can be studied with omics technologies and integrated with information from the functional physical structure using bio-imaging methods.

the functional physical structure using bio-imaging methods where e.g. the morphology, chlorophyll fluorescence, characterisation of pigments and functional studies are performed to better understand an organism (Figure 1.2B).

The link between the gene regulatory network and the functional physical structure (the double arrow in Figure 1.2A) is generally considered highly complex, with many pathways and pathway nodes interacting in what are often considered multifactorial processes. While it is undoubtedly flexible and adaptable to environmental constraints, the underlying links for a specific phenotype may turn out to be monofactorial, in particular in plants that can be grown under highly controlled conditions. The specific purpose of my thesis is to demonstrate that this can indeed be the case, and that reduction of complexity in systems biology may turn out to be the rule, rather than the exception, provided unbiased whole genome screens are performed to allow for unexpected factors. There are intriguing examples from other scientific disciplines as well, that indicate monofactorial connections. For instance, in host-pathogen interactions in human health, pathogens are likely to operate with very limited intervention interactions.[33] Thus, once the toolbox in Figure 1.2B for in depth analysis of the connections between the gene regulatory network and the functional physical structure is established, it can be applied in many areas of the life science, including human health.

In this PhD thesis, the ultimate goal is to understand the route for reducing the complexity of the different *Arabidopsis thaliana* mutants using metabolomics and to understand the underlying pathways of the phenotype of the mutants in a general framework.[34] Metabolomics is the study of all metabolites in an organism both qualitatively and quantitatively to get a clear metabolic picture of a living organism under specific conditions.[35-37] The metabolome is most closely related to the phenotype of a plant since metabolites are the end products of cellular processes.[35] Metabolomics is used to study development under normal and abiotic conditions (temperature, light, salt)[38] and biotic stress conditions (fungal, insects)[39,40], safety assessment of genetically modified crops[41], speed up crop improvements[42], effect of fruit storage[43] and the detection of food fraud.[44,45]

1.3 Analytical techniques in metabolomics

To study the metabolic profile of a plant, mass spectrometry (MS) or liquid-state nuclear magnetic resonance (NMR) spectrometry are the most common techniques in metabolomics. Both techniques have their own advantages and limitations as shown in Table 1.1. NMR spectrometry is a method which is non-destructive, with a high reproducibility and allows to quantify metabolites. On the other hand, while MS is more sensitive, which allows to detect more metabolites in a sample, it needs different chromatography techniques such as gas chromatography (GC) or liquid chromatography (LC) for different classes of metabolites.[42,46]

For both techniques, extraction of metabolites from the sample is necessary to obtain the metabolic profile. The drawback of this extraction is that it is not only time-consuming, but also that metabolites might be lost or degraded during extraction.[51] One way to eliminate the extraction procedure is to use high-resolution magic angle spinning (HR-MAS) NMR which allows using of intact tissue samples.[46,52,53]

Table 1.1: The advantages and limitations of NMR spectrometry and mass spectrometry for metabolic profiling.[47-50]

	NMR spectrometry	Mass spectrometry
Sensitivity	Low sensitivity, but can be improved with higher field strength and cryo- or microprobes.	High sensitivity, can reach the detection limit of attomolar (10^{-18}) concentrations.
Sample measurement	In one measurement with a detectable concentration can be detected.	Need chromatography techniques for different classes of metabolites.
Sample recovery	Non-destructive technique Several analyses can be performed on the same extracted sample.	Destructive technique.
Reproducibility	Very high.	Moderate.
Quantification	Absolute quantitation of metabolites possible by adding one standard with known concentration.	Quantification is possible with authentic standards, which are not available for newly identified compounds. Also, ionization efficiencies, ion suppression and matrix effects have influences on the concentration.
Targeted or untargeted approach	Untargeted approach.	Untargeted and targeted approach, but mainly used for targeted analysis.

1.4 Theoretical background of HR-MAS NMR

An NMR experiment can be described with a nuclear spin Hamiltonian:

$$H = H_{CS} + H_D^{IS} + H_D^{II} \tag{1.1}$$

Here,

$$H_{CS} = \{\sigma_{iso}\gamma B_0 + \tfrac{1}{2}\delta[3\cos^2(\theta) - 1 - \eta\sin^2(\theta)\cos(2\phi)]\}I_z \tag{1.2}$$

represents the chemical shift anisotropy interaction of the nuclei with the electronic environment,

$$H_D^{IS} = -\frac{\mu_0}{4\pi}\hbar\sum_i\sum_j\frac{\gamma^I\gamma^S}{r_{ij}^3}\frac{1}{2}(3\cos^2(\theta_{ij}) - 1)2I_z^iS_z^j \tag{1.3}$$

is the heteronuclear dipolar coupling between two different nuclear species I and S, and

$$H_D^{II} = -\frac{\mu_0}{4\pi}\hbar\sum_i\sum_j\frac{\gamma^2}{r_{ij}^3}\frac{1}{2}(3\cos^2(\theta_{ij}) - 1)(3I_z^iI_z^j - \mathbf{I}^i\cdot\mathbf{I}^j) \tag{1.4}$$

is the homonuclear dipolar coupling.[53-55]

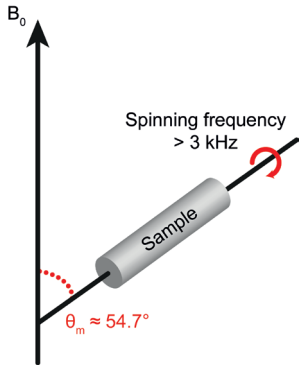


Figure 1.3: HR-MAS setup where the sample is rotated with high frequency (> 3 kHz) tiled by the magic angle θ_m with respect to the magnetic field B_0.

Here, σ_{iso} is the isotropic value, γ the gyromagnetic ratio and η is the asymmetry parameter. For the heteronuclear and homonuclear dipolar coupling, r_{ij} is the distance between the nuclei i and j and θ_{ij} is the angle between r_{ij} and the z-axis. The I spin is the abundant spin and S is the rare spins.

All three interaction terms depend on $\frac{1}{2}(3\cos^2(\theta)-1)$, where θ is the polar angle that describes the orientation of the magnetic field B_0 in the principle axis frame of the chemical shift tensor or dipolar interaction tensor. With HR-MAS NMR, the solid sample is rapidly rotated at the magic angle $\theta_m = 54.7°$. The angular dependences of the spin Hamiltonian will be averaged to zero over the sample and the broadening will be effectively removed (Figure 1.3). Although the anisotropic interactions produce spinning sidebands, these are suppressed when spinning at high frequencies (> 3 kHz), and the spectra will have narrow signals.

HR-MAS NMR is a combination of solid- and liquid-state NMR techniques, which can obtain spectra with similar resolution as spectra from liquid-state NMR experiments but make use of semi-solid samples with restricted molecular mobility.[56] Semi-solid samples, like biological tissues, can be used without extraction steps using this technique. In HR-MAS NMR, the effect of hetero- and homonuclear dipolar coupling is minimized at a frequency of a few kHz, while rigid solid samples need spinning frequencies of 20 – 50 kHz.

1.5 HR-MAS NMR-based workflow

In my PhD thesis, I propose an HR-MAS NMR-based workflow and apply the workflow to *Arabidopsis thaliana* mutants. The HR-MAS NMR-based workflow is shown in Figure 1.4. The workflow starts with the harvesting of the leaves from the *Arabidopsis thaliana* plant for preparation of a sample in the rotor, followed by performing the HR-MAS NMR experiments. The pulse sequences which can be used in metabolomics are described in section 1.6. The data are pre-processed and reduced by bucketing (section 1.7). Multivariate analysis is executed to determine biomarkers for the mutants in three steps: detection of outliers, investigation of the variation between different samples and selection of potential biomarker candidates (section 1.8). Finally, the biomarkers will be quantified and biological

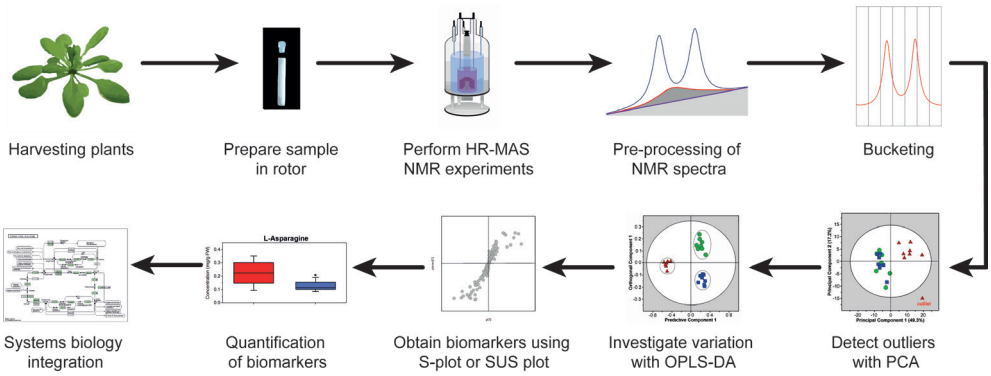


Figure 1.4: A typical high-resolution magic angle spinning (HR-MAS) NMR-based workflow

interpretation will be performed in a comprehensive systems biology approach by using available information from the literature. This workflow is based on a recently established liquid-state NMR approach[57], and the advantage of using HR-MAS NMR on leaves is that experiments can be performed genuinely *in vivo*, which will be illustrated with selected plant metabolomics applications (section 1.9).

1.6 Pulse sequences used in metabolomics

A complementary set of pulse sequences is used in NMR-based metabolomics to identify and quantify metabolites. One-dimensional spectra are used mostly to quantify metabolites. The mostly used pulse sequences are the one-dimensional ^{1}H NOESY (nuclear overhauser effect spectroscopy) with water presaturation and the ^{1}H CPMG (Carr-Purcell-Meiboom-Gill) sequence. NOESY spectra provide a complete and quantitative profile of the observed metabolites with the suppression of the water peak without an effect on the intensity of the other peaks.[58-60] CPMG is a pulse sequence which removes the broad signals from macromolecules, like proteins and lipids.[58,61]

In one-dimensional NMR spectra, signals from different metabolites strongly overlap. A way to solve this is to use two-dimensional NMR experiments. ^{1}H homonuclear correlation experiments are commonly used for identification. COSY (correlation spectroscopy) identifies spin-spin coupling of protons[58,61] and TOCSY (total correlation spectroscopy) provides information about the correlation between all protons in metabolites.[59,61] Another experiment is the ^{1}H *J*-resolved where the effect of chemical shift and *J*-coupling is separated into two independent directions.[62]

With the identification of new metabolites, it is sometimes helpful to make use of ^{1}H-^{13}C heteronuclear correlation experiments. These experiments provide information about the coupling between a proton and a carbon.[59,61] HSQC (heteronuclear single-quantum correlation) gives input about the correlation between a proton and a carbon which are separated by one bond. in addition, HMBC (heteronuclear multiple-bond correlation) gives information about the correlation over multiple bonds.[63]

Nearly all pulse sequences used for liquid samples in conventional high-resolution NMR

are also applicable for HR-MAS NMR. Out of the many pulse sequences that are available, three pulse sequences have been used for this work as shown in Figure 1.5. A CPMG pulse sequence with water suppression is used to obtain one-dimensional spectra (Figure 1.5A). The pulse sequence starts with a relaxation delay (d1) followed by a 90° pulse. Then there is looped n times over a spin echo delay (d20) followed by a 180° pulse and the FID signal is reordered. The CPMG pulse sequence is chosen because it minimizes the signal from lipids.[53] To confirm assignment, a magnitude-mode gradient-selected two-dimensional ^{1}H-^{1}H COSY and ^{1}H *J*-resolved pulse sequences (Figure 1.5B-C) can be used. The COSY pulse sequences start with water presaturation and a relaxation delay (d1) followed by a 90° pulse. After an increment delay (d0), a gradient pulse is executed, then a 90° pulse is applied and a gradient pulse at the same time followed by data acquisition. COSY spectra are very useful to resolve overlapping signals in the NMR spectra, especially in the aromatic region (6.0 – 8.0 ppm).[61] The long acquisition time of the COSY pulse sequence is a disadvantage. On the other hand, *J*-resolved NMR can acquire a spectrum within 20 minutes. The pulse sequence of the *J*-resolved experiment is shown in Figure 1.5C. In this sequence, after a relaxation delay d1, a 90° pulsed is applied followed by an increment delay d0 and a 180° pulse followed by a second increment delay and the FID signal is recorded.[61,62]

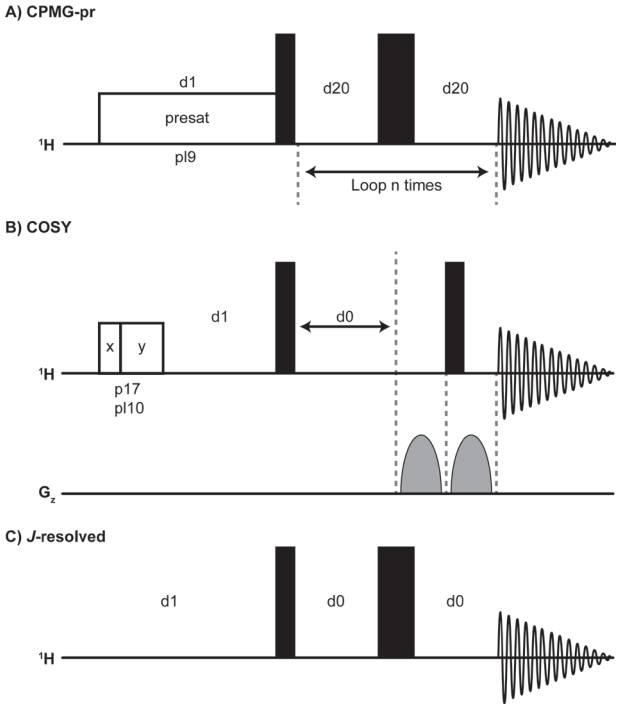


Figure 1.5: Pulse sequences for (A) the CPMG ("cpmgpr1d"), COSY ("cosygpppqf") (B) and J-resolved ("jresqf") (C) experiments. Narrow and wide black bars indicate 90° and 180° pulses, respectively. d0 is the increment delay, d1 is the relaxation delay and d20 is the fixed echo time to allow elimination of J modulation effects. pl9 is the power level for presaturation. pl10 is the power level of TOCSY-spinlock and p17 is the trim pulse at pl10.

1.7 Pre-processing of one-dimensional HR-MAS NMR spectra

Prior to multivariate analysis and quantification, raw spectra need to be pre-processed. Incorrect pre-processing can lead to spurious results.[64,65] For one-dimensional ^{1}H NMR spectra, pre-processing involves alignment, baseline correction, bucketing, normalization and scaling.

Spectral alignment

NMR resonances can be shifted due to several factors such as changes in pH, temperature, salt concentration and inhomogeneous magnetic fields. This can give rise to variations between spectra collected from the same sample species. To solve this problem, standard chemical shifts can be used for the metabolites, and the spectra can be aligned to the standard to construct a dataset for multivariate analysis.[61,65] A more elegant, unbiased protocol to align the spectra is by using an internal shift reference, since this leaves the relative shifts unaffected. This is done by adding a reference compound with a known chemical shift with the sample. Most often 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (TSP) or 4-4-dimethyl-4-silapentane-1-sulfonic acid (DDS) is used as a reference compound. Both compounds have a methyl resonance with 0 ppm chemical shift relative to tetramethylsilane (TMS), the standard reference across the entire field of ^{1}H NMR spectroscopy.[64,65]

Baseline correction

The NMR responses of metabolites are superimposed on a broad background that does not contribute any signal of interest but affects the multivariate analysis and impedes quantification of metabolites. Polynomial-fitting of the regions in between the NMR signals is used to perform automated baseline correction.[65] After baseline correction, the spectra are truncated to have only signals from metabolites. The region between 0.1 and 8 ppm is used for further analysis. Although water suppression is employed during acquisition, a weak remaining water signal can interfere with the multivariate data analysis and the region of the water peak around 4.8 ppm is also removed.[64-66]

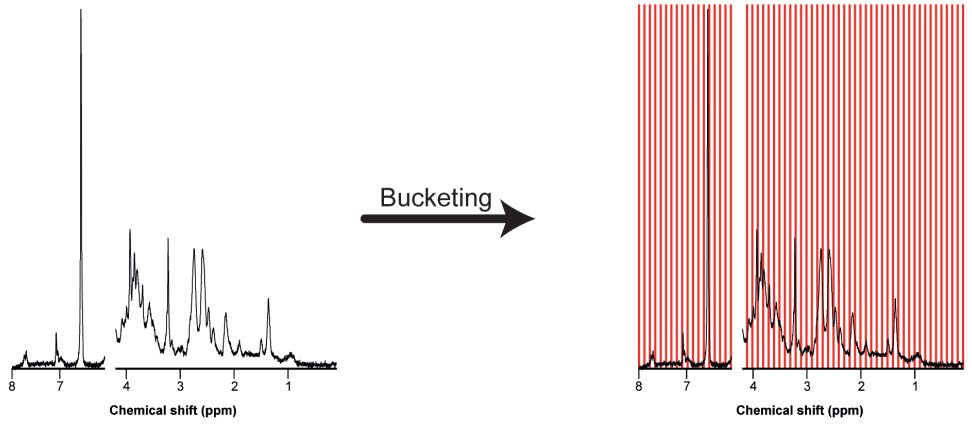


Figure 1.6: Truncated NMR spectrum before and after bucketing into equally spaced buckets of 0.04 ppm width. Bucketing allows for moderate shift averaging at the expense of resolution and provides a matrix for further processing (Figure 1.7).

Bucketing

The truncated NMR spectra typically have around 22.000 data points. It is common to reduce the resolution of the data by bucketing, also known as binning.[64-66] The spectral intensities are summed over equally spaced buckets of 0.04 ppm width (Figure 1.6). This procedure averages minor variations in chemical shift and reduces the amount of data for the multivariate analysis.[64-66] After bucketing, an $i \times j$ data matrix X is obtained with on the rows the different samples, while the columns represent the chemical shifts. The elements of the matrix contain the intensity of the bins, i.e. the signal at the different shifts for each sample.

Normalization

Biological differences between preparations, for instance different weight or dilution, result in different concentrations of specific metabolites. Normalization methods aim to remove such systematic errors.[64,65] A standard method is to normalize the individual samples (i.e. rows) of the bucket matrix X according to

$$x_{ij} = \frac{x_{ij}}{\sum_1^j x_i} \tag{1.5}$$

This is illustrated in Figure 1.7 for a simple case of three samples.

Scaling

Since higher concentration metabolites generally also exhibit the strongest variation, scaling of the columns is necessary to avoid only the selection of the most abundant metabolites in the multivariate analysis.[65]

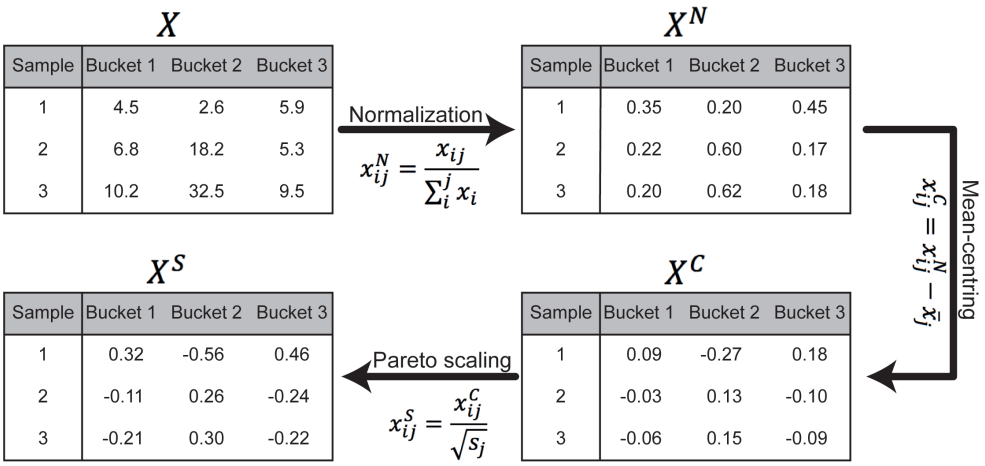


Figure 1.7: Every data point in the bucket matrix X ($i \times j$) is normalized. Every column of the normalized data matrix is mean centred and scaled using the Pareto method. The obtained data matrix X^S is used for multivariate analysis. x_{ij} is an element located in the ith row and the jth column. $\bar{x}_j$ and s_j are respectively the mean and standard deviation of the values of the jth column.

In this thesis, the Pareto scaling method has been used (Figure 1.7).[65] The first step in Pareto scaling is mean-centring of the samples, where the high-concentration and low-concentration metabolites are converted to values which vary around zero by subtracting the mean values from the columns. Pareto scaling uses the square root of the standard deviation from the columns as a scaling factor. This provides data X^S which is closely related to the real data to study the covariance of the data matrix in multivariate analysis.[65-67]

1.8 Multivariate analysis

Multivariate analysis considers multiple variables simultaneously to identify patterns in the data corresponding to signal patterns from metabolites.[65-68] These generally contain more than one proton, and their signals are therefore spread over several buckets. First, unsupervised methods, methods with no assumption of any prior knowledge, are used to explore the data, find outliers and group the data.[65-68] In this thesis, unsupervised principal component analysis (PCA) is performed, where an orthogonal transformation is used to convert the set of correlated intensities (Bucket 1, Bucket 2, ..., Bucket n) with coordinates x_{ij}^S for the samples into a set of linearly uncorrelated intensities called principal components $(\mathbf{PC_1}, \mathbf{PC_2}, ..., \mathbf{PC_n})$. PCA operates with two mathematical constraints, largest possible variance and orthogonality. The first principal component $\mathbf{PC_1}$ has the largest possible variance under the linear transformation. The subsequent vectors $\mathbf{PC}_i$ are orthogonal to the preceding components and each has the highest possible variance in their coordinates under the constraints of the prior vectors $(\mathbf{PC_1}, ..., \mathbf{PC_{i-1}})$.[65,67,69] PCA is converting the correlated X^S into an uncorrelated orthogonal basis set of vector components $(\mathbf{PC_1}, \mathbf{PC_2}, ..., \mathbf{PC_n})$, containing the scores, the new coordinates of the samples. Scores are represented in a two-dimensional score plot where each point represents a single sample on two principal component coordinates (Figure 1.8A). The transformation matrix that provides the connection with the data after preprocessing is named the loadings and describe how the old bucket intensities are linearly combined to the principal components and indicate which buckets have the most influences on the principal component and can

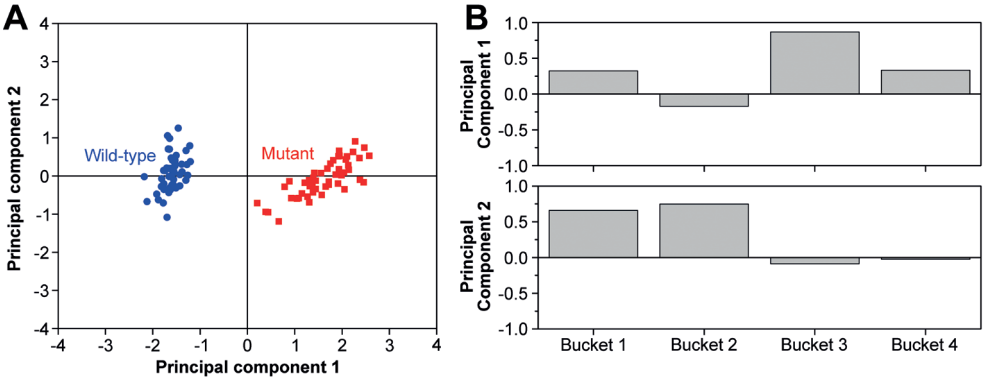


Figure 1.8: PCA score (A) and loading plot (B) of a data set including 50 wild-type samples and 50 mutant samples and 4 buckets for every sample. The score plot shows a clear separation between wild-type and mutant. The PCA loading plot shows that bucket 3 has the most influences on the first principal component and bucket 1 and 2 has the most influences on the second principal component.

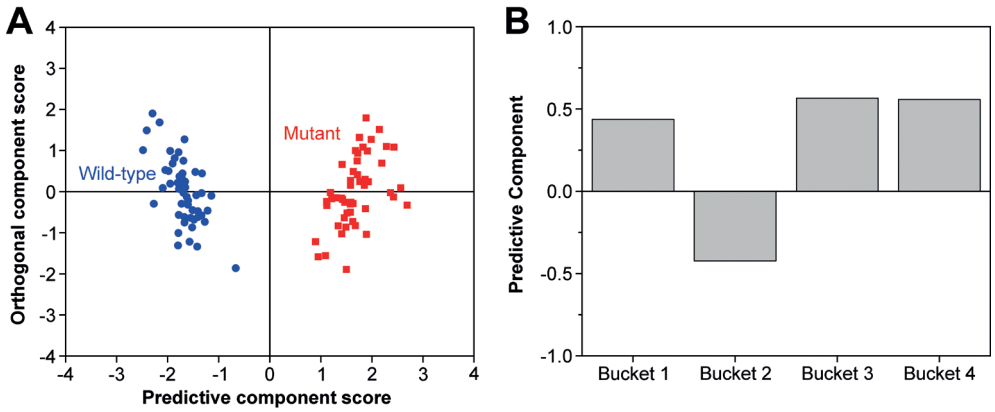


Figure 1.9: OPLS-DA score (A) and loading plot (B) for the same data set as described in Figure 1.8. In the OPLS-DA model, there is also clear separation between wild-type and mutant in the score plots. The loading plot shows that bucket 3 and bucket 4 have the most influence on the between-class variation.

be represented into a loading plot (Figure 1.8B).[67,70-72] The next step is to use databases, like the biological magnetic resonance bank (BMRB) and the human metabolome database (HMDB), to identify the metabolites corresponding to these buckets and to perform further downstream systems biology analyses.[73]

Supervised methods is used to cluster the data and to determine biomarkers by following how clusters of buckets representing a specific metabolite change between e.g. wild-type and a specific mutant. The model is applied with a priori knowledge of sample classes. Supervised methods can, therefore, be used to mark the separation between two or more sample classes at the level of individual metabolites.[65,67,69] In this thesis, orthogonal partial least squares discriminant analysis (OPLS-DA) has been utilized. OPLS-DA is a multiple regression method which uses the pre-processed data matrix X^S and a newly defined vector $\mathbf{y}$ with the value 0 for the wild-type samples and 1 for the mutant.[74] The data matrix X^S is split into a part correlated to $\mathbf{y}$, also named the predictive component (X_p^S), and another part that is uncorrelated to $\mathbf{y}$, also called the orthogonal component (X_o^S)[66,69,71-74]:

$$X^S = X_p^S + X_o^S = T_p P_p^T + T_o P_o^T \qquad (1.6)$$

In the formula above, T represents the score matrix and P the loading matrix which can be represented respectively in a score plot and a loading plot (Figure 1.9). The loading plot of the predictive component represents the between-class variation, *i.e.* wild-type vs mutants, and indicates which buckets have the strongest impact on the variation. The metabolites corresponding to these buckets are identified using metabolome databases.[73] Description of the technical details of the OPLS-DA procedure falls outside the scope of this thesis, and the reader is referred to Trygg and Wold.[75]

1.9 Applications of HR-MAS in plant metabolomics

HR-MAS NMR combined with multivariate analysis can be a powerful tool to study plant metabolomics. However, HR-MAS NMR is not used very often in plant biology. Over the last decades, ±30 publications report on HR-MAS NMR-based metabolomics studies in plants. These publications are summarized in Table 1.2.

The HR-MAS NMR-based metabolomics studies in plants have been used for wide range of applications. The influences of biotic and abiotic stress on the metabolic profile has been predominantly studied by in HR-MAS NMR. Examples are the influences of drought on plants[76-79] and the effect of fungicides or pesticides on the plants.[80-85] Metabolomics can help in understanding developmental processes, like fruit ripening. The metabolic profile throughout the ripening process is studied in mango[86] and tomato[87]. The impact of storage time on the metabolic profile is studied on Golden Delicious apples.[88] HR-MAS NMR-based metabolomics can also be used to study the metabolic profile of specific cell type to understand the plant better. Mucci *et al*. studied different tissues of lemons and citrons to understand the similarities and the differences between these two fruits.[89] In addition, it is possible to use metabolic profiling to characterize newly discovered plants[90] or mutants of plants.[91-93] The geographical origin of sweet peppers[94], garlic[95] and cocoa beans[96] has been investigated with HR-MAS NMR. The original geographical origin of some food products has been certified, for example with Protected Geographical Indications (PGIs). HR-MAS NMR-based metabolomics is a useful tool for these certified products to avoid fraud.[44] Examples are the cherry tomatoes of Pachino[97,98], Interdonato lemon of Messina[98,99] and tomatoes from Almería.[100] HR-MAS NMR can also be used to determine different classes or cultivars of plants. This is helpful when only one class has a medical application as in the case of *Trichilia catigua*[101] or *Withania somnifera*.[102] It also can help to distinguish different cultivars of apples[103], melons[104], rice[105] and persimmons.[106]

Table 1.2: Summary of the publications of metabolomics studying using high-resolution magic angle spinning NMR

Plant	Research objective	Magnetic field strength (MHz)	Pulse sequences	Multivariate models
Influences of biotic or abiotic stress				
Winter wheat (*Triticum aestivum*)[76]	Evaluate the influences of different drought treatments	400	1D	PCA
Soybean[77]	Determine the influences of water deficiency	600	CPMG, NOESY	PLS-DA
Jatropha curcas[78]	Determine the impacts of pruning procedures and water management	400	Zg	-

Ribes nigrum[79]	Determine the effect of seasonal asymmetric warming	600	CPMG, HSQC	PCA
Jatropha curcas[80]	Studying the effect of *Jatropha* mosaic virus	400	NOESY, CPMG, COSY	-
Pear (*Pyrus communis*) and Quince (*Cydonia oblonga*)[81]	Study the effect of humic acid on the morphogenesis	400	CPMG, COSY, TOCSY, HSQC	PCA
Lettuce (*Lactuca sativa*)[82]	Influences of the fungicide mancozeb on the leaves at different growth stages	800	NOESY, TOCSY, HSQC	PCA, PLS-DA
Tomato (*Solanum lycopersicum*)[83]	Study the influences of 6-pentyl-2H-pyran-2-one and harzianic acid on the leaves	400	CPMG, COSY, TOCSY, *J*-res, HSQC, HMBC	PCA
Maize (*Zea mays*)[84]	Determine the toxic effects on maize root tips of organo-chlorine pesticides	600	CPMG	OPLS-DA
Maize (*Zea mays*)[85]	Determine the effect of mineral or compost fertilization and inoculation with arbuscular mycorrhizal fungi	400	CPMG, COSY, TOCSY, *J*-res, HSQC, HMBC	PCA
Study the ripening and storage of fruits				
Mango fruit (*Mangifera indica*)[86]	Studying the metabolic profile of mango pulp during ripening	400	^{1}H 1D, TOCSY, *J*-res	-
Tomato (*Solanum lycopersicum*)[87]	Studying different tissues of the tomato during fruit ripening	500	NOESY, TOCSY, HMQC	PCA
Golden Delicious apples[88]	Determine the impact of storage time and production systems	500	NOESY, COSY, TOCSY	PCA, PLS-DA
Studying different cell types of plants				
Lemon (*Citrus limon*) and Citron (*Citrus medica*)[89]	The metabolic profile of different parts of the lemon and citron are studied	400	^{1}H, CPMG, COSY, TOCSY, HSQC	-
Characterizing of plant				
Crocus sativus[90]	Establish the main metabolites present in *C. sativus* petals	400	^{1}H, COSY, TOCSY, HSQC, HMBC	-
Understanding transgenic plants				
Poplar tree (*Populus tremula*)[91]	Studying the time- and growth-related metabolic profile of PttMYB76 and wild-type poplar tree	500	CPMG	PCA, PLS-DA

Common bean (*Phaseolus vulgaris*)[92]	Distinction between wild-type and transgenic common beans	500	CPMG	PCA
"Swingle" citrumelo[93]	Evaluate the metabolic profile of non- and transgenic citrumelo	500	^{1}H, HSQC, TOCSY	PCA, PLS-DA
Geographical origin of plants				
Sweet peppers (*Capsicum annum*)[94]	Discriminate peppers according to their geographical origin	400	NOESY, 1D ^{13}C, TOCSY	PLS-DA
Garlic (*Allium sativum*)[95]	Characterisation of two varieties cropped in different regions	400	NOESY, ^{13}C, TOCSY, HMQC	PLS-DA
Cocoa beans[96]	Assess the geographical origins of fermented and dried cocoa beans	400	^{1}H	PCA, PLS-DA, OPLS-DA
Cherry tomatoes of Pachino[97]	Determine the major metabolites present in cherry tomatoes of Pachino	700	^{1}H	PCA
PGI Cherry Tomato of Pachino, PGI Inter-donato Lemon of Messina, Red Garlic of Nubia[98]	Identifing and quantifying metabolites from 3 typical food products of the Mediterranean diet	700	^{1}H	PCA
PGI Interdonato Lemon of Messina[99]	Determine metabolites unique for PGI Interdonato Lemon of Messina	700	^{1}H, COSY, TOCSY, HSQC	-
Tomato (*Lycopersicon esculentum*)[100]	Establish differences between commercially available varieties	500	NOESY, HSQC	PCA
Distinguish between different cultivars				
Trichilia catigua[101]	Classification of commercial samples of Catuaba	400	CPMG	PCA, HCA
Withania somnifera[102]	Evaluate metabolic profile of different chemotypes	800	CPMG, COSY, HSQC	PCA
Apples[103]	Discriminate 3 different apple cultivars by their metabolic profile	500	NOESY, COSY, TOCSY	PCA, PLS-DA
Melon (*Cucumis melo*)[104]	Quantification of sugars and compare two varieties	400	^{1}H	-
Rice (*Oryza sativa*)[105]	Determine the metabolic variation of diverse rice cultivars	700	CPMG, TOCSY, HSQC, STOCSY	PCA, OPLS-DA
Persimmon (*Diospyros kaki*)[106]	Follow the metabolic changes during development of different cultivars	400	NOESY	PCA

1.10 Outline of this thesis

The purpose of this thesis is to understand the mechanisms of enhanced growth characteristics the phenotypically engineered *Arabidopsis* mutants by establishing a non-invasive metabolic approach using HR-MAS NMR combined with multivariate analysis.

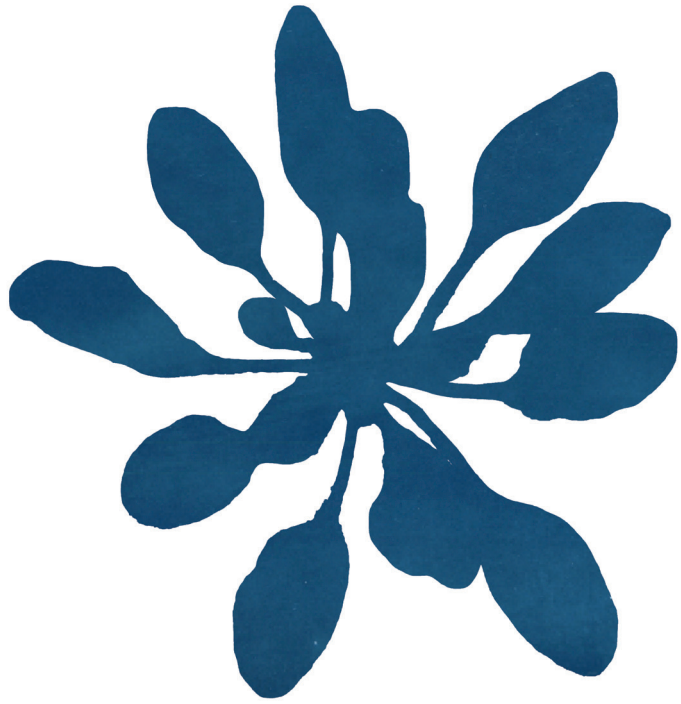
In **Chapter 2**, the *in vivo* metabolic profile has been established directly from the *Arabidopsis thaliana* leaves using HR-MAS NMR throughout the circadian cycle to reveal primary metabolites and their functional periodicity of the circadian rhythm. In **Chapter 3**, HR-MAS NMR is used to obtain the metabolic profile from *Arabidopsis thaliana* mutants with enhanced growth characteristics at the middle of the light period. Combined with multivariate analysis, this approach suggests that both mutants have a diminished defence response leading to an altered growth-defence trade-off. Physiological functions like growth and defence are regulated by the circadian rhythm. In **Chapter 4**, the circadian rhythm of the identified metabolites is studied in both *Arabidopsis* mutant with enhanced growth characteristics. This shows that the circadian rhythm of the metabolites was not affected, only the levels of the metabolites differ throughout the circadian cycle. **Chapter 5** provides a general discussion of the work presented in this thesis and future prospects.

1.11 References

[1] UNFCCC, Report No. FCCC/CP/L.9/Rev.1 2015.
[2] UNGA Resolution ARES 2015.
[3] United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects: the 2017 Revision, Key Findings and Advance Tables, 2017.
[4] International Energy Agency, Tracking Clean Energy Progress, 2017.
[5] International Energy Agency, Technology Roadmap; Delivering Sustainable Bioenergy, 2017.
[6] C. Mba et al., Agriculture & Food Security 2012, 1, 7–17.
[7] A. Kumar et al., OMICS 2015, 19, 581–601.
[8] A. S. Edison et al., Metabolites 2016, 6, 8.
[9] M. Koornneef, D. Meinke, Plant J. 2010, 61, 909–921.
[10] U. Krämer, eLife Sciences 2015, 4, e06100.
[11] X. Zhang et al., Nat. Protocols 2006, 1, 641–646.
[12] J. M. Van Norman, P. N. Benfey, Wiley Interdiscip Rev Syst Biol Med 2009, 1, 372–379.
[13] F. Breseghello, A. S. G. Coelho, J. Agric. Food Chem. 2013, 61, 8277–8286.
[14] N. van Tol, B. J. van der Zaal, Plant Science 2014, 225, 58–67.
[15] T. K. Mohanta et al., Genes (Basel) 2017, 8, 399.
[16] B. I. Laufer, S. M. Singh, Epigenetics & Chromatin 2015 8:1 2015, 8, 34.
[17] T. Gaj et al., Trends in Biotechnology 2013, 31, 397–405.
[18] A. Beltran et al., Assay Drug Dev Technol 2006, 4, 317–331.
[19] T. Sera, Advanced Drug Delivery Reviews 2009, 61, 513–526.
[20] D. J. Segal et al., Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 2758–2763.
[21] B. I. Lindhout et al., The Plant Journal 2006, 48, 475–483.
[22] A. Bent, in Agrobacterium Protocols (Ed.: K. Wang), Humana Press, Totowa, 2006, pp. 87–104.
[23] N. van Tol et al., PLoS One 2017, 12, e0174236.
[24] N. van Tol et al., Sci. Rep. 2017, 7, 3314.
[25] N. van Tol et al., Plant Cell Environ 2016, DOI 10.1111/pce.12805.
[26] N. van Tol, Phenotypic Engineering of Photosynthesis Related Traits in *Arabidopsis Thaliana* Using Genome Interrogation, 2016.
[27] C. Eynck et al., in Biofuel Crop Sustainability, John Wiley & Sons, Ltd, 2013, pp. 165–204.

[28] H. I. Gomes, Environmental Technology Reviews 2012, 1, 59–66.
[29] J. S. Yuan et al., Trends in Plant Science 2008, 13, 165–171.
[30] T. Ogura, W. Busch, Annu. Rev. Cell Dev. Biol. 2016, 32, 103–126.
[31] B. P. Sheth, V. S. Thaker, Planta 2014, 240, 33–54.
[32] G. W. Bassel et al., Plant Cell 2012, 24, 3859–3875.
[33] A. Dix et al., Clin Microbiol Infect 2016, 22, 600–606.
[34] E. K. F. Chan et al., PLoS Genet 2010, 6, e1001198.
[35] O. Fiehn, Plant Mol Biol 2002, 48, 155–171.
[36] H. K. Kim et al., Nat Protoc 2010, 5, 536–549.
[37] N. Schauer, A. Fernie, Trends in Plant Science 2006, 11, 508–516.
[38] I. Ahuja et al., Trends in Plant Science 2010, 15, 664–674.
[39] J. W. Allwood et al., Physiologia Plantarum 2008, 132, 117–135.
[40] J. J. Jansen et al., Metabolomics 2008, 5, 150.
[41] C. Simó et al., Int. J. Mol. Sci. 2014, 15, 18941–18966.
[42] R. Kumar et al., Front. Plant Sci. 2017, 8, 443.
[43] S. Brizzolara et al., Postharvest Biology and Technology 2017, 127, 76–87.
[44] E. Cubero-Leon et al., Food Chemistry 2018, 239, 760–770.
[45] D. I. Ellis et al., Innovation in food science - Foodomics technologies 2016, 10, 7–15.
[46] A.-H. M. Emwas, in Metabonomics, Humana Press, New York, NY, 2015, pp. 161–193.
[47] A.-H. M. Emwas et al., Metabolomics 2013, 9, 1048–1072.
[48] A.-H. M. Emwas, in Metabonomics: Methods and Protocols (Ed.: J.T. Bjerrum), Springer New York, New York, NY, 2015, pp. 161–193.
[49] J. L. Markley et al., Current Opinion in Biotechnology 2017, 43, 34–40.
[50] F. Matsuda, Mass Spectrometry 2016, 5, S0052.
[51] T. S. Maier et al., Plant Methods 2010, 6, 6–6.
[52] J. Z. Hu, Metabolomics : open access 2016, 6, e147.
[53] O. Beckonert et al., Nat Protoc 2010, 5, 1019–1032.
[54] A. Alia et al., Photosynth Res 2009, 102, 415–425.
[55] P. Mazzei, A. Piccolo, Chemical and Biological Technologies in Agriculture 2017, 4, 11.
[56] M. Vermathen et al., Chimia (Aarau) 2012, 66, 747–751.
[57] C. Deborde et al., Progress in Nuclear Magnetic Resonance Spectroscopy 2017, 102-103, 61–97.
[58] J. Kruk et al., Appl Magn Reson 2017, 48, 1–21.
[59] B. Elena-Herrmann, in NMR-Based Metabolomics, Royal Society of Chemistry, Cambridge, 2018, pp. 22–38.
[60] A. Le Guennec et al., Anal. Chem. 2017, 89, 8582–8588.
[61] A. C. Dona et al., Comput Struct Biotechnol J 2016, 14, 135–153.
[62] C. Ludwig, M. R. Viant, Phytochem. Anal. 2009, 21, 22–32.
[63] T. De Meyer et al., Anal Bioanal Chem 2010, 398, 1781–1790.
[64] L. R. Euceda et al., Scand J Clin Lab Invest 2015, 75, 193–203.
[65] A. Smolinska et al., Analytica Chimica Acta 2012, 750, 82–97.
[66] K. H. Liland, TrAC Trends in Analytical Chemistry 2011, 30, 827–841.
[67] B. Worley, R. Powers, CMB 2012, 1, 92–107.
[68] X. Qi et al., Eds., Plant Metabolomics, Springer Netherlands, Dordrecht, 2015.
[69] I. T. Jolliffe, J. Cadima, Phil. Trans. R. Soc. A 2016, 374, 20150202.
[70] J. Trygg et al., J. Proteome Res. 2007, 6, 469–479.
[71] J. Trygg, S. Wold, J. Chemometrics 2002, 16, 119–128.
[72] M. Bylesjö et al., J. Chemometrics 2006, 20, 341–351.
[73] J. J. Ellinger et al., CMB 2013, 1, 10.2174/2213235X11301010028.
[74] S. Wiklund et al., Anal. Chem. 2008, 80, 115–122.

[75] J. Trygg, S. Wold, J. Chemometrics 2002, 16, 119–128.
[76] H. Winning et al., Journal of Experimental Botany 2008, 60, 291–300.
[77] I. D. Coutinho et al., Phytochem. Anal. 2017, 28, 529–540.
[78] O. N. A. Santos et al., Industrial Crops and Products 2017, 109, 918–922.
[79] M. Pagter et al., J. Agric. Food Chem. 2017, 65, 10123–10130.
[80] O. P. Sidhu et al., Planta 2010, 232, 85–93.
[81] G. Marino et al., J. Agric. Food Chem. 2013, 61, 4979–4987.
[82] S. I. Pereira et al., Food Chemistry 2014, 154, 291–298.
[83] P. Mazzei et al., J. Agric. Food Chem. 2016, 64, 3538–3545.
[84] C. Blondel et al., Environ Pollut 2016, 214, 539–548.
[85] P. Mazzei et al., J. Agric. Food Chem. 2018, acs.jafc.7b04340.
[86] A. M. Gil et al., J. Agric. Food Chem. 2000, 48, 1524–1536.
[87] E. M. S. Pérez et al., Food Chemistry 2010, 122, 877–887.
[88] M. Vermathen et al., Food Chemistry 2017, 233, 391–400.
[89] A. Mucci et al., Food Chemistry 2013, 141, 3167–3176.
[90] V. Righi et al., J. Agric. Food Chem. 2015, 63, 8439–8444.
[91] S. Wiklund et al., Plant Biotechnology Journal 2005, 3, 353–362.
[92] R. Choze et al., Food Chemistry 2013, 141, 2841–2847.
[93] C. S. de Oliveira et al., Magn. Reson. Chem. 2014, 52, 422–429.
[94] M. Ritota et al., J. Agric. Food Chem. 2010, 58, 9675–9684.
[95] M. Ritota et al., Food Chemistry 2012, 135, 684–693.
[96] A. Marseglia et al., Food Research International 2016, 85, 273–281.
[97] D. Mallamace et al., Physica A 2014, 401, 112–117.
[98] C. Corsaro et al., J Anal Methods Chem 2015, 2015, 175696.
[99] N. Cicero et al., Nat Prod Res 2015, 29, 1894–1902.
[100] E. M. S. Pérez et al., FRIN 2011, 44, 3212–3221.
[101] C. Daolio et al., Phytochem. Anal. 2008, 19, 218–228.
[102] S. K. Bharti et al., Magn. Reson. Chem. 2011, 49, 659–667.
[103] M. Vermathen et al., J. Agric. Food Chem. 2011, 59, 12784–12793.
[104] T. Delgado-Goñi et al., Planta 2013, 238, 397–413.
[105] E.-H. Song et al., J. Agric. Food Chem. 2016, 64, 3009–3016.
[106] A. D. D. C. Santos et al., Food Chemistry 2018, 239, 511–519.



Metabolic profiling of intact *Arabidopsis thaliana* leaves during circadian cycle using ^{1}H high-resolution magic angle spinning NMR

D. Augustijn
U. Roy
R. van Schadewijk
H. J. M. de Groot
A. Alia

Published in PLoS ONE 11(9): e0163258

2.1 Abstract

Arabidopsis thaliana is the most widely used model organism for research in plant biology. While significant advances in understanding plant growth and development have been made by focusing on the molecular genetics of *Arabidopsis thaliana*, extracting and understanding the functional framework of metabolism is challenging, both from a technical perspective due to losses and modification during extraction of metabolites from the leaves, and from the biological perspective, due to random variation obscuring how well the function is performed. The purpose of this work is to establish the *in vivo* metabolic profile directly from the *Arabidopsis thaliana* leaves without metabolite extraction, to reduce the complexity of the results by multivariate analysis, and to unravel the mitigation of cellular complexity by predominant functional periodicity. To achieve this, we use the circadian cycle that strongly influences metabolic and physiological processes and exerts control over the photosynthetic machinery. High-resolution magic angle spinning (HR-MAS) NMR was applied to obtain the metabolic profile directly from intact *Arabidopsis* leaves. Combining one- and two-dimensional ^{1}H HR-MAS NMR allowed the identification of several metabolites including sugars and amino acids in intact leaves. Multivariate analysis on HR-MAS NMR spectra of leaves throughout the circadian cycle revealed modules of primary metabolites with significant and consistent variations of their molecular components at different time-points of the circadian cycle. Since robust photosynthetic performance in plants relies on the functional periodicity of the circadian rhythm, our results show that HR-MAS NMR promises to be an important non-invasive method that can be used for metabolomics of the *Arabidopsis thaliana* mutants with altered physiology and photosynthetic efficiency.

2.2 Introduction

As a model organism, *Arabidopsis thaliana* plays a central role in understanding biological functions across plant species and in characterizing phenotypes associated with genetic mutations.[1] Significant advances in understanding plant growth and development have been made by focusing on the molecular genetics of *Arabidopsis thaliana*. Several high-throughput technologies to produce information on the transcriptome, metabolome, proteome, interactome and other omics datasets are available.[2,3] However, understanding the functional framework of metabolism in native state in leaves poses a major challenge for all metabolomics approaches. Many approaches, including mass spectrometry as well as NMR methods, require labour-intensive extraction of plant metabolites which can cause biases resulting from differential extraction efficiencies and from the loss of volatile metabolites.[4,5] Extraction methods also cause the loss of molecular information regarding specific associations within and between polymeric structural plant components.

Understanding the functional framework of metabolism is also challenging from biological perspective due to random variations obscuring how well the function is performed. It has been argued that periodicity in a biological system such as circadian rhythms can provide robustness that helps to tolerate the random variations.[6] Many biological systems rely on functional periodicity, as is evidenced by abnormal or chaotic behaviour when functional periodicity is lost. In plants, the circadian cycle strongly influences metabolic and physiological processes.[7-10] The endogenous biological clock allows plants to anticipate on daily changes in the environment, such as the onset of dawn and dusk, providing them with an adaptive

advantage. It has been shown that growth, productivity and competitive advantage in plants are enhanced by matching the circadian cycle with the external light/dark cycle.[7] The internal clock also regulates physiological processes, including photoperiodic induction of flowering, hypocotyl elongation, cotyledon movement and stomatal opening.[8-10] Previous studies reported large diurnal changes in the expression of several genes in *Arabidopsis thaliana.*[11,12] Diurnal changes of few soluble metabolites have also been reported in extracts of *Arabidopsis* leaves.[13] The current understanding of diurnal changes in metabolites has been based on destructive analysis of individual components.[11-13] These in vitro results may not faithfully reflect the native structural and conformational information. Examining the rhythmic pattern of metabolites directly in the intact *Arabidopsis thaliana* leaves without any extraction during circadian cycle would be important to understand the functional framework of metabolism in the native state.

High-resolution magic angle spinning (HR-MAS) NMR offers a fast and sensitive method to study molecules in intact samples *in situ* and *in vivo*. HR-MAS NMR is viewed as a hybrid technique between solution state NMR and solid-state NMR. Similar to solid-state NMR, the use of magic angle spinning (MAS) effectively removes spectral line broadening resulting from magnetic susceptibility, homonuclear dipolar interactions and chemical shift anisotropy. When the sample is spinning along the magic angle of $\theta = 54.7°$ with respect to the static magnetic field (B_0), line broadening effects are reduced to zero because the $\frac{1}{2}(3\cos^2(\theta)-1)$ part of the Hamiltonian disappears.[14] Thus, HR-MAS NMR yields narrow lines in heterogeneous samples such as tissue or whole cells. At the same time, it retains the advantages of low power levels and deuterium locking in classical solution NMR experiments for obtaining good stability, resolution and overall performance of the NMR experiment. Similar to solution NMR, HR-MAS NMR involves direct polarization transfer and not cross polarization transfer (CPMAS) between ^{1}H and other nuclei (^{13}C or ^{15}N), thus differentiating it from CPMAS experiments on true solids.

The application of HR-MAS NMR has been earlier reported for studying chemotype variations in frozen leaf and root samples of *Withania somnifera*[15], to monitor alterations in metabolite profile of *Jatropha curcas* during virus infection[16] and to detect metabolites in extracts from *Arabidopsis thaliana.*[17] In addition, the application of HR-MAS NMR for metabolite monitoring has been reported for Italian sweet pepper,[18] Italian garlic,[19] citrumelo,[20] and for tree species such as poplar[21] and *Euglena*.[22] To our knowledge, HR-MAS NMR was not yet applied to intact fresh leaves of *Arabidopsis thaliana*.

The objective of the present study was to establish the metabolic profile directly from the *Arabidopsis thaliana* leaves without metabolite extraction using HR-MAS NMR spectroscopy and to study functional framework of metabolism by following metabolic rhythm throughout the light/dark cycle. Our results demonstrate that HR-MAS NMR on intact *Arabidopsis thaliana* leaves represents a novel platform that could provide important in *vivo* information of regular metabolic network against which altered metabolic profile due to stress, infection or mutation can be assessed.

2.3 Materials and Methods

Material

Seeds of wild-type *Arabidopsis thaliana* plants of ecotype Colombia O (Col-0) were incubated in Petri dishes on wet filtration paper and transferred to small tubes filled with soil and sand mixture (Holland Potgrond). To synchronise germination, the tubes were kept in complete darkness at 277 K for 72 hours. Germinated seeds were transferred to the greenhouse and maintained at 293 K under a 12 hours light (200 µmol $m^{-2}s^{-1}$) and 12 hours dark regime for 4 weeks, at which time flowering had not commenced.

Sample collection and preparation for NMR analysis

Typically, 8 - 10 replicate samples of intact rosette leaves were collected at 15 time-points during the entire photoperiod (Figure 2.1). For observing the internal metabolite rhythm during free running conditions, the plants were transferred to continuous dark for 48 hours. Intact rosette leaves were then collected at 15 time-points within 24 hours. A single leaf was rolled and inserted into a 4 mm Zirconium Oxide rotor. 10 µL of deuterated phosphate buffer (100 mM, pH 6) containing 0.1% (w/v) 3-trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) was added as a lock solvent and NMR reference, respectively. The rotor was placed immediately inside the NMR spectrometer. For each time-point, eight replicates were measured.

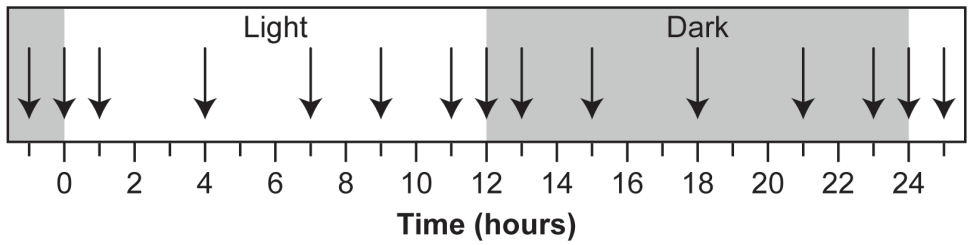


Figure 2.1: Time-points of harvesting of non-flowering Arabidopsis rosette leaves during the circadian cycle during growth stage 3.70 – 3.90.

^{1}H high-resolution magic angle spinning (HR-MAS) NMR spectroscopy

All experiments were carried out with a Bruker DMX 400 MHz NMR spectrometer operating at a proton resonance frequency of 399.427 MHz and equipped with a 4 mm HR-MAS dual inverse $^{1}H/^{13}C$ probe with a magic angle gradient. Data were collected with a spinning frequency of 4 kHz. A temperature of 277 K was used to avoid any tissue degradation during data acquisition. The temperature was stabilized with a Bruker BVT3000 control unit.

One-dimensional ^{1}H HR-MAS NMR spectra were recorded using the rotor synchronized Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence with water suppression.[23] Each one-dimensional spectrum was acquired in 16k data points applying a spectral width of 8000 Hz. The number of averages was 256, with 8 dummy scans. A constant receiver gain of 1024, an acquisition time of 2 seconds and a relaxation delay of 2 seconds were used. Since NMR measurements were done on intact tissue, the relaxation delay was set to a small value to remove nascent short T_2 components due to the presence of lipids. Spectra were processed by applying an exponential window function corresponding to a line broadening

of 1 Hz and data were zero-filled prior to Fourier transformation. ^{1}H HR-MAS NMR spectra of plant tissue were phased manually and automatically baseline corrected using TOPSPIN 2.1 (Bruker Analytische Messtechnik, Germany).

To confirm the assignments, two-dimensional homonuclear correlation spectroscopy (^{1}H-^{1}H COSY) was performed using Bruker's standard pulse program library. The parameters used for COSY were: 2048 data points were collected in the t_2 domain over the spectral width of 4k Hz, 512 t_1 increments were collected with 64 transients, relaxation delay 1 sec, acquisition time 116 msec. The data were zero-filled to 512 data points and were weighted with a sine bell window function in both dimensions prior to Fourier transformation. Two-dimensional *J*-resolved spectrum was measured using pulse sequence ("jresqfpr"), from the Bruker pulse program library. Representative *J*-resolved spectrum is shown in Supplementary figure 2.2.

Quantification of metabolites

NMR data analysis was performed using MestReNova software version 10.0.1–14719 (Mestrelab Research S.L. Spain). The concentrations of the various metabolites in the spectra of intact leaf were determined by comparing the integral peak intensity of the metabolite of interest with that of the TSP peak, after correcting for the number of contributing protons and for tissue weight. All statistical analysis (t-tests and ANOVAs) of the NMR quantification results were performed with OriginPro v. 9 (Northampton, USA). F-values were calculated, and F-values larger than 2.8 ($p < 0.05$) were considered significant.

Multivariate analysis

Multivariate analysis of primary metabolites in the spectra was performed using the Bruker software package AMIX (version 3.8.6). The one-dimensional CPMG spectra, collected from leaf samples at 1, 7, 12 and 23 hours (see Figure 2.1), were subdivided in the range between 0.3 and 9 ppm into buckets of 0.04 ppm (total 218 buckets), using Bruker AMIX software (Version 3.8.7, Bruker GmbH). The region of 4.20 – 6.00 ppm was excluded from the analysis to remove the water signal. To compensate for the differences in the overall metabolite concentration between individual samples, the data obtained were mean centred, scaled by the Pareto method and then normalized by dividing each integral of the segment by the total area of the spectrum.[24] The resulting data matrix was exported into Microsoft Office Excel (Microsoft Corporation, USA). This was then further imported into SIMCA software (Umetrics AB) for multivariate statistical analysis.

2.4 Results and Discussion

Identification of metabolites

Since the metabolites in intact cells of leaves may differ dramatically in their abundance, size, location and relative mobility, a rotor-synchronized Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence coupled with water suppression was used to improve sensitivity and to generate a better baseline by removing backgrounds, resulting from superimposition of molecules in low abundance and/or with restricted mobility. A representative one-dimensional ^{1}H HR-MAS NMR spectrum obtained directly from the intact leaf of *Arabidopsis thaliana* at t = 7 hours is shown in Figure 2.2. Peak assignment was performed according to earlier literature and the Biological Magnetic Resonance Data Bank (BMRB).[25,26] A list of identified

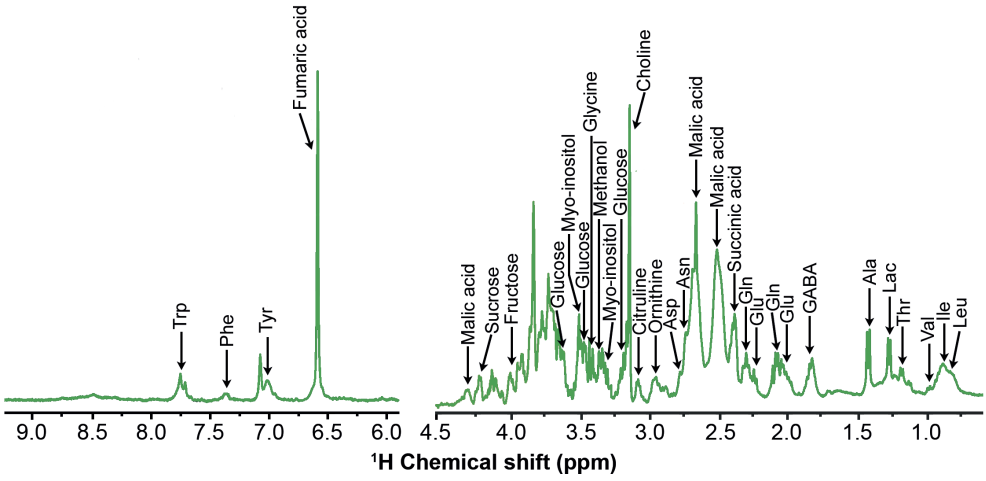


Figure 2.2: A representative one-dimensional ^{1}H HR-MAS NMR spectrum, obtained from the intact leaf of Arabidopsis thaliana at t = 7 hours, showing resonance assignment of several metabolites.

metabolites is given in Supplementary table 2.1.

The ^{1}H HR-MAS NMR spectrum could be divided into three major regions. The high-field region (0.0 – 3.0 ppm) was rich in amino acids, the mid-field region (3.0 – 5.5 ppm) contained sugars and the down-field region (5.5 – 10.0 ppm) was dominated by aromatic compounds. In the high-field region, several metabolites were identified including L-alanine, L-threonine, lactic acid, γ-aminobutyric acid (GABA), L-glutamic acid and malic acid. Signals from mid-field region showed diverse sugars. Signals from fumaric acid, L-tyrosine, L-tryptophan and L-phenylalanine were observed in the down-field region. The results were corroborated by two-dimensional COSY spectra (Supplementary figure 2.1) and *J*-resolved spectra (Supplementary figure 2.2) to resolve the complexity of overlapping and interfering spectral regions to allow for correct identification of metabolites. Detailed assignment of sugar region in the two-dimensional COSY spectrum is shown in Supplementary figure 2.3.

Among several metabolite signals, twelve primary metabolites were quantified by integrating the distinct characteristic signals of each metabolite with respect to the intensity of the nine protons of TSP on the fresh weight basis. These metabolites include organic acids (fumaric acid, malic acid and lactic acid), sugars (glucose, fructose), amino acids (L-glutamic acid, L-alanine, L-phenylalanine, L-tyrosine, L-aspartic acid and GABA), and precursor of membrane phospholipids (e.g. choline).

Characterisation of metabolites throughout the circadian cycle

The circadian cycle strongly influences many plant metabolic and physiological processes.[7-10] Previous studies reported large diurnal changes in the expression of many genes in *Arabidopsis thaliana*.[11-13] To understand the functional framework of metabolism in native state in leaves during circadian cycle, we examined metabolites at different time-points of the circadian cycle directly in the intact leaves. Figure 2.3 shows significant and

consistent rhythmic pattern of several metabolites during circadian cycle. The stacked plots at different time-points are shown in Supplementary figure 2.4.

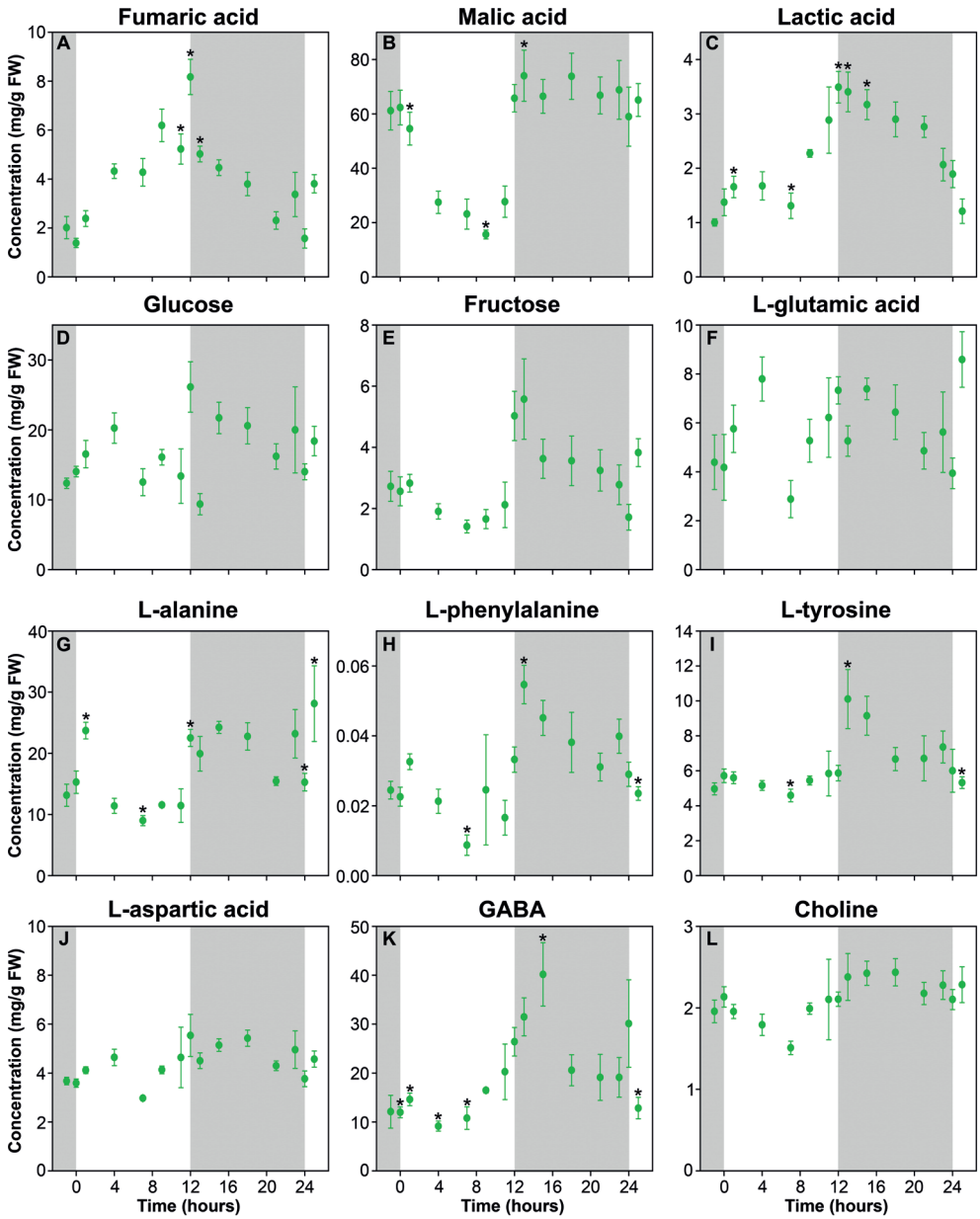


Figure 2.3: Changes in the levels of metabolites during circadian cycle in leaves of Arabidopsis thaliana. The whole intact rosette leaf was harvested at different time-points during light and dark period. The results are given as the mean ± standard error ($p < 0.05$).*

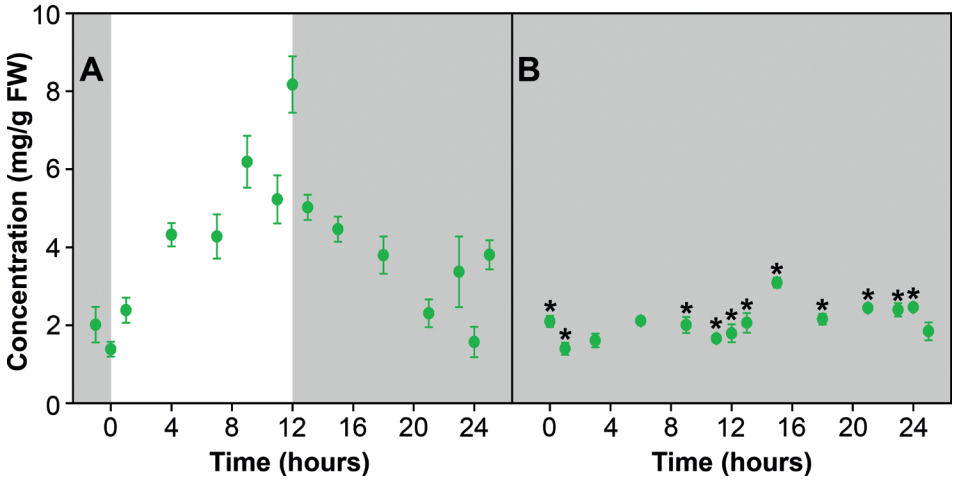


Figure 2.4: Changes in the levels of fumaric acid in the leaves of Arabidopsis thaliana. Fumaric acid was measured in intact leaves at different time-points during 12h light/12h dark period (A) and during continuous dark period (B). Results are mean of 4 replicates ± standard error ($p < 0.05$).*

Fumaric acid participates in a multiplicity of pathways in plant metabolism, however, its function as carbon stores in C3 plants has not been deeply addressed. While in C3 plants, the major photoassimilates are starch and soluble sugars, in some of the C3 plants, including *Arabidopsis*, fumaric acid is considered to be one of the major forms of fixed carbon.[27] Previous studies have indicated that similar to starch and soluble sugars, fumaric acid can be metabolized to yield energy and carbon skeletons for production of other compounds. Figure 2.3A shows that fumaric acid concentrations increases during the day, reaching maximum at the end of the light period and then started decreasing and reached its minimum level at the end of the dark period. This observation is consistent with previous study showing high level of fumaric acid during light period measured in the extract of *Arabidopsis* leaves by GC-MS.[28] Interestingly, the concentration of fumaric acid dropped to a steady level in *Arabidopsis* shifted to extended dark (Figure 2.4B). A possible explanation is that the formation and the degradation rate of fumaric acid may be equal during continuous dark. It is also possible that fumaric acid is transported out of the leaves during growth in continuous dark.[27]

Malic acid is another carbon storage molecule which participates in various pathways in plant metabolism and also plays an important role in CAM and C4 photosynthesis.[28] Malic acid concentration showed a decreasing pattern during the light period, while it increased during dark period and remained high during the dark period (Figure 2.3B). This is in contrast to earlier studies where diurnal malic acid changes assayed by GC-MS in *Arabidopsis* grown in a 16h light/8h night regime showed a high level of malic acid at the end of the day and declined during night time.[29] This difference could be attributed to differences in light/dark regime[13] used in previous study as well as extraction methods used which cause the release of malic acid from different compartments in the plant cells. Malic acid is dominantly compartmentalised in vacuole.[30,31] In our study, the signals of $^{5'}CH_2$ and ^{2}CH of malic acid were slightly shifted (observed at 2.5 ppm and 4.35 ppm, respectively) as compared to the signals of malic acid in water at pH 7.0 ($^{5'}CH_2$ and ^{2}CH at 2.35 and 4.29, respectively).[32]

Similar shifts of resonances of malic acid have been observed in leaves in earlier HR-MAS NMR studies[15], which may be due to the influence of native acidic environment of the vacuole, where excess of malic acid is stored. Isolating malic acid by extraction methods not only remove the compartmentalised information but also may cause potential chemical changes or degradation due to absence of native environment. Loss of compartmental information may mask critical details on its ability to perform *in vivo* rhythmic function. Thus, non-invasive measurement of malic acid by HR-MAS NMR as shown in this study has clear advantage in monitoring diurnal changes in malic acid in its native environment.

Figure 2.3C shows rhythmic changes in lactic acid concentration. The concentration of lactic acid was highest at the end of the light period and dropped during the dark period. The role of lactic acid in *Arabidopsis* leaves is still unclear and the diurnal changes in lactic acid have not been studied so far. However, it has been recognised in previous studies that insertion of electrons coming from lactic acid into the respiratory electron transport chain in light-dependent.[33]

The sugar derivatives, glucose and fructose levels are reported to be influenced by light/dark period.[11,12] Figure 2.3D and Figure 2.3E show the concentration of glucose and fructose throughout the light/dark cycle. An increasing trend was observed in the glucose concentration during the light period, which declined during the dark period. The concentration of fructose in the leaves first declined during the light period followed by a rise to a maximum amount at the end of the light period. Subsequently, it declined during the dark period. During the light period, photosynthetic carbon fixation takes place in the leaves and photosynthate is stored primarily in the form of sugars. These sugars are then utilized during the dark period.[11] The circadian rhythm of sugars is known as an important player in the global regulation of diurnal gene expression.[11,34,35] The low concentration of sugars at the beginning of the light period ensures the repression of the PSEUDO-RESPONSE REGULATORY7 (PRR7) promoter, leading to the activation of the morning-expressed CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) genes in the central loop.[34,35]

The L-glutamic acid level was elevated during the light period and dropped during the dark period as shown in Figure 2.3F. The L-glutamic acid level is known to be dependent on available nitrate.[36] Nitrate is taken up from the soil and is converted to ammonium in the leaves or roots and reduced through the glutamine synthetase/glutamate synthase (GS/GOGAT) pathway, to L-glutamic acid or L-glutamine.[36,37] Recent studies have shown that nitrate assimilation is stimulated by light[11], which may be responsible for the elevation of L-glutamic acid as seen in the present study (Figure 2.3F). The α-amino group of L-glutamic acid can be transferred to other amino acids via various aminotransferases.[37] Thus, the rhythmic pattern of L-glutamic acid may influence the level of other amino acids throughout the light/dark cycle. Figure 2.3G-J illustrate that L-alanine, L-phenylalanine, L-tyrosine as well as L-aspartic acid show a clear rhythmic pattern during the light/dark cycle. L-glutamic acid is also known to be converted into the ubiquitous non-protein amino acid γ-aminobutyric acid (GABA), an import player in plant carbon metabolism, mainly because it can bypass two steps in the TCA cycle.[38,39] This is especially useful when the plants are grown under carbon-limited conditions. Figure 2.3K shows that GABA concentration in *Arabidopsis* leaves

was enhanced during the light period and decreased during later in the dark period. This phenomenon was also observed in an earlier study of Fahnenstich *et al*. using GC-MS.[29]

The level of choline has also been monitored during the light/dark period (Figure 2.3L). The concentration of choline decreased during the light period and increased during the dark period (Figure 2.3L). Choline is an essential metabolite in plants needed to synthesize membrane phospholipids. It is widely distributed and occurs in relatively high concentrations in many plant tissues as free choline, lipid choline (phosphatidyl choline), and sometimes as water-soluble bound choline (phosphocholine and glycerophosphocholine).[40,41] Up to now, the pattern of choline during a light/dark period has not been studied in *Arabidopsis*. It is known, however, that the time of the day influences the membrane lipid composition of *Arabidopsis* leaves.[42,43] Our results indicate that periodic behaviour of choline may be closely linked with periodic changes in membrane lipids composition.

Multivariate analysis of ^{1}H HR-MAS NMR spectra of Arabidopsis leaves

Multivariate analysis has proven to be useful in metabolomics. It can explain the variance within a dataset, helps with identification of biologically relevant spectral features or to identify outliers.[44] The two most popular methods for multivariate analysis are unsupervised principal component analysis (PCA) and supervised orthogonal partial least squares discriminant analysis (OPLS-DA). These methods give both a score matrix and a loading matrix, with score matrix showing the relation between observations, while the loading matrix gives the individual contribution of each parameter, which is a peak in the case of NMR spectra.[44-48] OPLS-DA was used on the ^{1}H HR-MAS NMR spectra from the leaves collected at the beginning of the light period (1 hour), at the middle of the light period (7 hours), at the start of the dark period (13 hours) and at the end of the dark period (23 hours) for covering important time period range. With the predictive components 1, 2 and 3, 74% of the total variance can be explained (Figure 2.5). Even though not completely separated, a clear clustering could be observed from the four time-points.

The loading plot of all buckets containing peaks which have been assigned in Supplementary

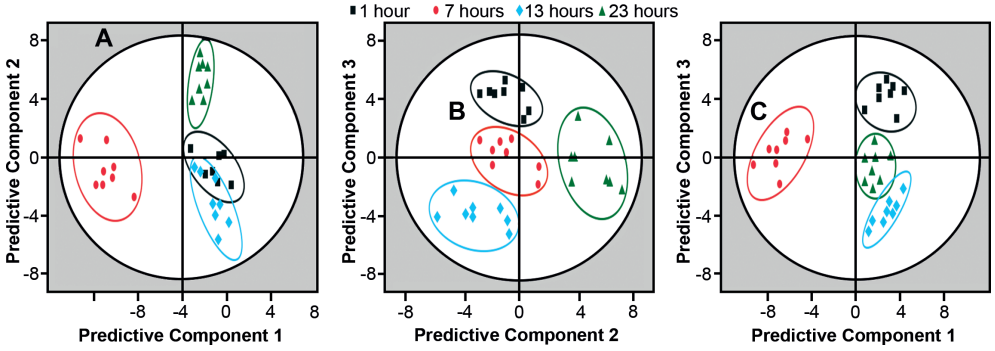


Figure 2.5: Score plots of OPLS-DA analysis of the HR-MAS NMR spectra showing the variable responsible for the separation among samples collected at different times during the circadian cycle from Arabidopsis Col-0.

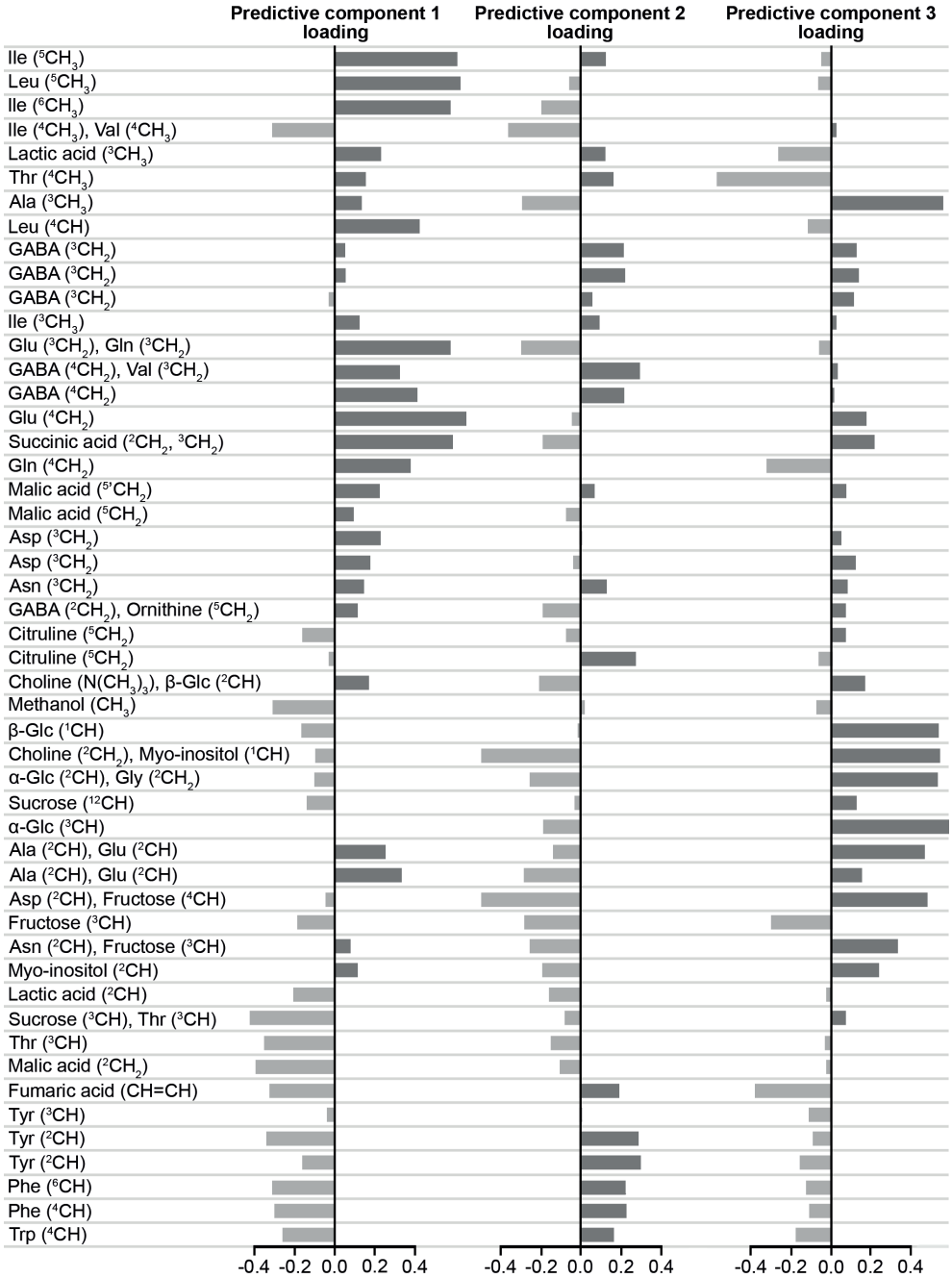


Figure 2.6: Loading plots of predictive component 1, 2 and 3 for all buckets of the OPLS-DA model (Figure 2.5) containing assigned peaks (Supplementary table 2.1). The metabolites on the vertical axis are arranged in accordance with their chemical shift order in the spectrum.

table 2.1 is shown in Figure 2.6. Examination of the loadings coming from predictive component 1 shows that the separation between the time-points arises due to positive loading of GABA, L-aspartic acid, malic acid, L-glutamine, L-glutamic acid, L-alanine and negative loading corresponding to L-phenylalanine, L-tyrosine, fumaric acid and lactic acid. The predictive components 2 loading showed positive loading of L-phenylalanine, L-tyrosine, glucose, GABA, L-aspartic acid, malic acid, L-alanine and negative loading of fumaric acid, lactic acid, malic acid, L-glutamine and L-glutamic acid. The predictive components 3 loading showed positive loading of L-glutamine, L-glutamic acid, glucose, GABA and L-alanine and negative loading of L-phenylalanine, L-tyrosine, fumaric acid and lactic acid (Figure 2.6). This suggested that these metabolites differ in concentration throughout the light/dark cycle.

2.5 Conclusion

In vivo analysis of metabolic profile during circadian cycle, especially in intact leaves, is an unconquered frontier. The current understanding of metabolic profile in *Arabidopsis thaliana* has been based on analysis with destructive extraction procedures. The analytical methods that provide *in vivo* information could yield important and novel insights into the cellular complexity of metabolic pathways in plants. Multidimensional high-resolution magic angle spinning (HR-MAS) NMR has evolved to be a powerful technique in a variety of *in vivo* studies including intact cells. In this study, multidimensional HR-MAS NMR was successfully applied for the first time on intact *Arabidopsis thaliana* leaves to obtain the metabolic profile throughout the circadian cycle to unravel cellular complexity predominated by functional periodicity. Multivariate analysis revealed clear variations in primary metabolites at different time-points of the light/dark cycle. Knowledge of *in vivo* metabolic profile at different time-points in the circadian cycle will be important for choosing the particular time-point to compare the metabolites pattern in wild-type and different available mutants of *Arabidopsis thaliana* with altered physiology and metabolism.

2.6 References

[1] D. W. Meinke et al., Science 1998, 282, 662–682.
[2] J. M. Van Norman, P. N. Benfey, Wiley Interdiscip Rev Syst Biol Med 2009, 1, 372–379.
[3] A. R. Joyce, B. O. Palsson, Nat Rev Mol Cell Biol 2006, 7, 198–210.
[4] H. K. Kim et al., Nat Protoc 2010, 5, 536–549.
[5] K. A. Kaiser et al., Magn. Reson. Chem. 2009, 47 Suppl 1, S147–56.
[6] J. D. Thomas et al., Annu Rev Biophys Biomol Struct 2004, 33, 75–93.
[7] A. N. Dodd, Science 2005, 309, 630–633.
[8] A. Samach, G. Coupland, Bioessays 2000, 22, 38–47.
[9] S. Barak et al., Trends in Plant Science 2000, 5, 517–522.
[10] J. A. Kreps, S. A. Kay, Plant Cell 1997, 9, 1235–1244.
[11] O. E. Bläsing et al., Plant Cell 2005, 17, 3257–3281.
[12] Y. Gibon et al., Genome Biol. 2006, 7, R76.
[13] R. Sulpice et al., Molecular Plant 2014, 7, 137–155.
[14] O. Beckonert et al., Nat Protoc 2010, 5, 1019–1032.
[15] S. K. Bharti et al., Magn. Reson. Chem. 2011, 49, 659–667.
[16] O. P. Sidhu et al., Planta 2010, 232, 85–93.
[17] Y. Sekiyama et al., 2010, 82, 1643–1652.
[18] M. Ritota et al., J. Agric. Food Chem. 2010, 58, 9675–9684.

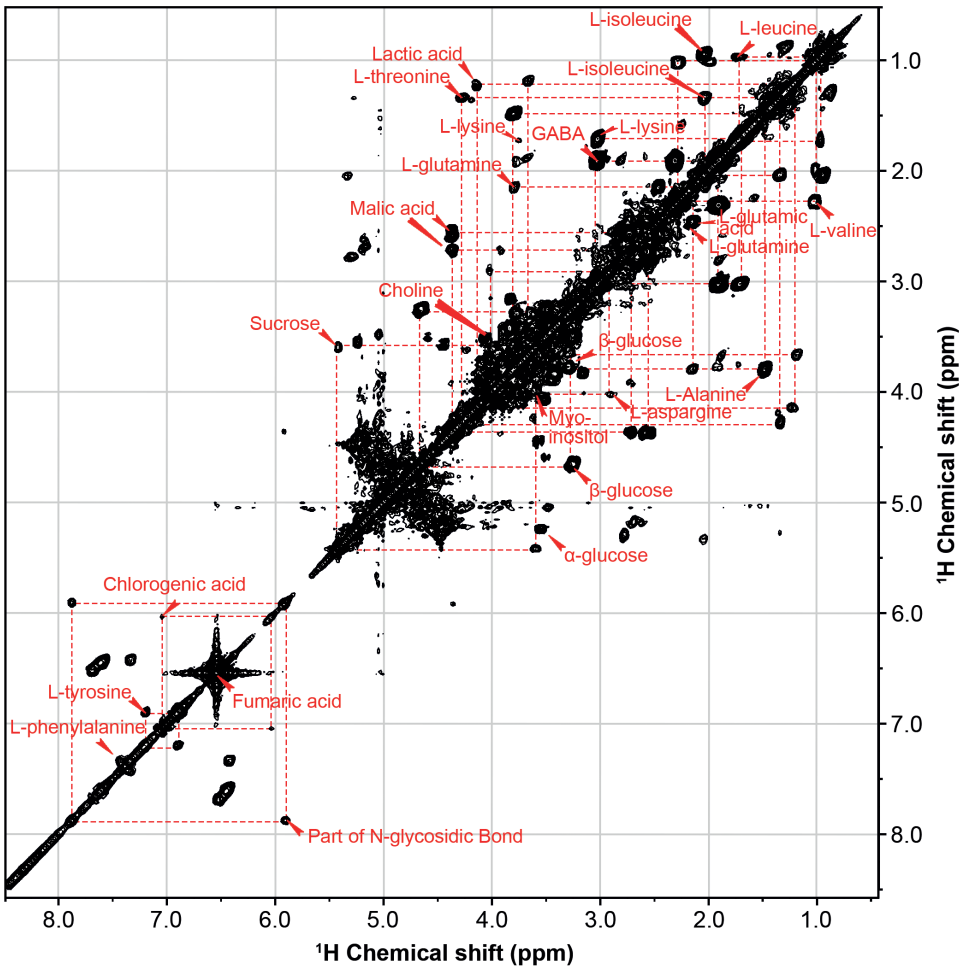
[19] M. Ritota et al., Food Chemistry 2012, 135, 684–693.
[20] C. S. de Oliveira et al., Magn. Reson. Chem. 2014, 52, 422–429.
[21] T. Mori et al., Sci Rep 2015, 5, 11848.
[22] T. Komatsu et al., Environ Sci Technol 2015, 49, 7056–7062.
[23] S. Meiboom, D. Gill, Review of Scientific Instruments 1958, 29, 688–691.
[24] M. Vermathen et al., J. Agric. Food Chem. 2011, 59, 12784–12793.
[25] V. Govindaraju et al., NMR Biomed 2000, 13, 129–153.
[26] E. L. Ulrich et al., Nucleic Acids Res 2008, 36, D402–8.
[27] D. W. Chia et al., Planta 2000, 211, 743–751.
[28] M. B. Zell et al., Plant Physiology 2010, 152, 1251–1262.
[29] H. Fahnenstich et al., Plant Physiology 2007, 145, 640–652.
[30] T. Tohge et al., Plant Physiology 2011, 157, 1469–1482.
[31] R. Gerhardt et al., Plant Physiology 1987, 83, 399–407.
[32] D. S. Wishart et al., Nucleic Acids Res 2013, 41, D801–7.
[33] P. Schertl, H.-P. Braun, Front. Plant Sci. 2014, 5, DOI 10.3389/fpls.2014.00163.
[34] M. J. Haydon et al., Nature 2013, 502, 689–692.
[35] A. N. Dodd et al., Front. Plant Sci. 2015, 6, DOI 10.3389/fpls.2015.00245.
[36] M. Stitt et al., Journal of Experimental Botany 2002, 53, 959–970.
[37] B. G. Forde, P. J. Lea, Journal of Experimental Botany 2007, 58, 2339–2358.
[38] S. Michaeli et al., Plant J. 2011, 67, 485–498.
[39] S. Michaeli, H. Fromm, Front. Plant Sci. 2015, 6, 43.
[40] Y. Li-Beisson et al., Arabidopsis Book 2013, 11, e0161.
[41] S. D. McNeil et al., PNAS 2001, 98, 10001–10005.
[42] S. Maatta et al., Front Plant Sci 2012, 3, 49.
[43] Y. Nakamura et al., 2014, 9, e29715.
[44] B. Worley, R. Powers, CMB 2012, 1, 92–107.
[45] Z. Ramadan et al., Talanta 2006, 68, 1683–1691.
[46] J. Bartel et al., Comput Struct Biotechnol J 2013, 4, e201301009–9.
[47] M. Bylesjö, in Methods in Molecular Biology (Ed.: J.T. Bjerrum), Springer New York, 2015, pp. 137–146.
[48] A. Alonso et al., Front Bioeng Biotechnol 2015, 3, 23.

2.7 Supplementary information

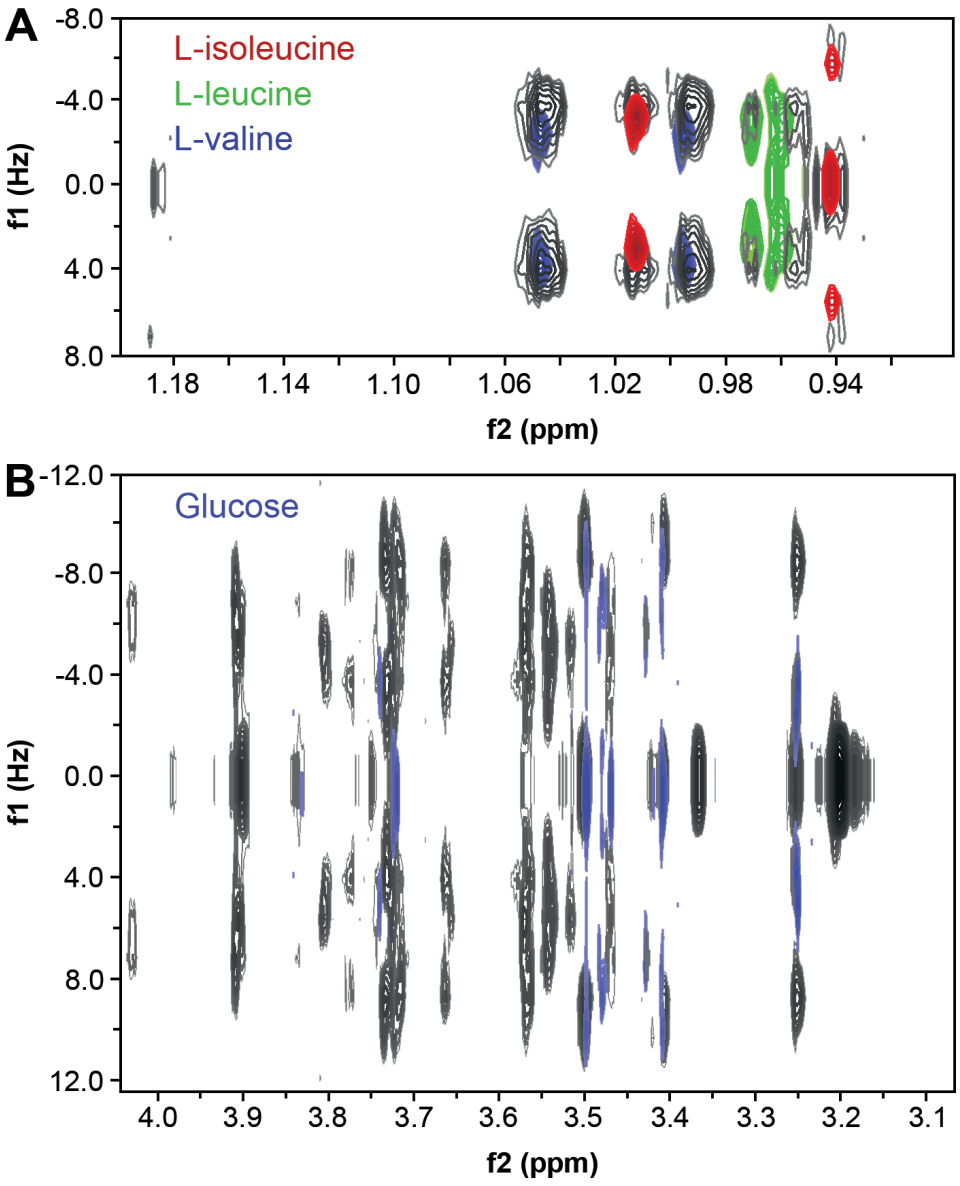
Supplementary table 2.1: ^{1}H Chemical shift assignment of metabolites in leaves of wild-type Arabidopsis thaliana [25,26]

Compound	Assignment	Chemical shift (ppm)	Multiplicity	Connectivity
L-alanine	^{2}CH $^{3}CH_3$	3.76 1.46	q d	2-3
L-aspartic Acid (Asp)	^{2}CH $^{3}CH_2$	3.90 2.80	dd dd	2-3
L-asparagine (Asn)	^{2}CH $^{3}CH_2$	4.00 2.80	dd m	2-3
Choline	$^{1}CH_2$ $^{2}CH_2$ $N(CH_3)_3$	4.05 3.50 3.22	m m s	1-2
Citrulline	$^{5}CH_2$	3.13	t	
Fructose	^{3}CH ^{4}CH	4.03 3.89	t dd	4-5
Fumaric Acid	CH=CH	6.60	s	
GABA	$^{4}CH_2$ $^{3}CH_2$ $^{2}CH_2$	2.28 1.89 3.00	t m t	4-3 3-2
α-D-glucose (Glc)	^{1}CH ^{2}CH ^{3}CH	5.22 3.52 3.70	d dd t	1-2 2-3
β-D-glucose (Glc)	^{1}CH ^{2}CH ^{3}CH	4.63 3.23 3.47	d dd t	1-2 2-3
L-glutamic acid (Glu)	^{2}CH $^{3}CH_2$ $^{4}CH_2$	3.74 2.12 2.35	dd m m	 2-3 3-4
L-glutamine (Gln)	$^{3}CH_2$ $^{4}CH_2$	2.11 2.43	m m	3-4
L-glycine	$^{2}CH_2$	3.54	s	
L-isoleucine (Ile)	$^{3}CH_3$ $^{4}CH_2$ $^{5}CH_3$ $^{6}CH_3$	1.97 1.25 0.92 0.98	m m t d	3-5 3-6
Lactic acid	^{2}CH $^{3}CH_3$	4.10 1.32	q d	2-3

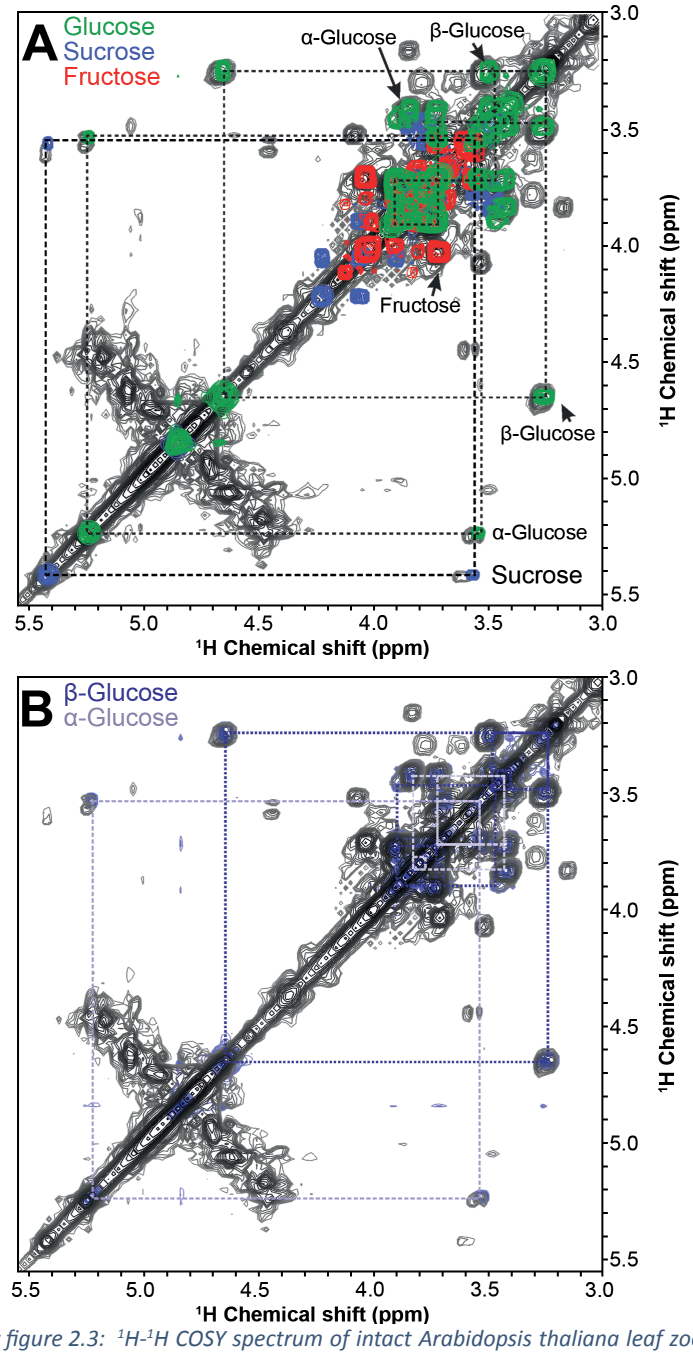
L-leucine (Leu)	^{4}CH $^{5}CH_3$	1.70 0.95	m t	4-5
Malic Acid	$^{2}CH_2$ $^{5}CH_2$ $^{5'}CH_2$	4.29 2.64 2.50	dd dd dd	 2-5 2-5'
Methanol	CH_3	3.35	s	
Myo-Inositol	^{1}CH ^{2}CH	3.52 4.05	dd t	1-2
Ornithine	$^{5}CH_2$	3.02	t	
L-phenylalanine (Phe)	^{4}CH ^{6}CH	7.36 7.32	m m	4-6
Succinic Acid	$^{2}CH_2$ $^{3}CH_2$	2.39 2.39	s s	
Sucrose	^{3}CH ^{7}CH ^{12}CH	4.22 5.42 3.59	d d dd	 7-12
L-threonine (Thr)	^{3}CH $^{4}CH_3$	4.24 1.32	m d	3-4
L-tryptophan (Trp)	^{4}CH	7.73	d	
L-tyrosine (Tyr)	^{2}CH ^{3}CH	7.19 6.89	m m	2-3
L-valine (Val)	$^{3}CH_2$ $^{4}CH_3$	2.26 1.03	m d	3-4



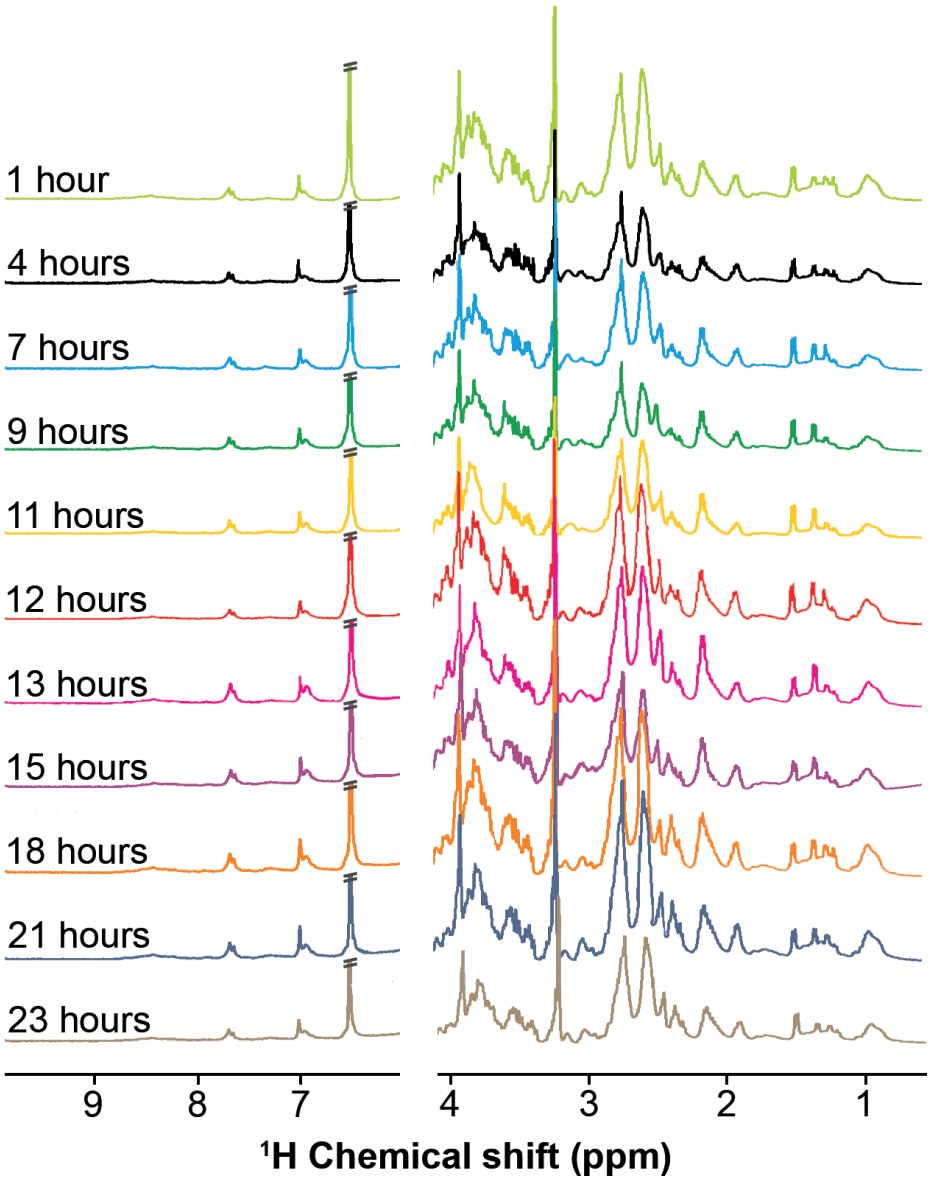
Supplementary figure 2.1: A representative two-dimensional HR-MAS ^{1}H-^{1}H COSY spectrum obtained from the intact leaf of Arabidopsis thaliana.



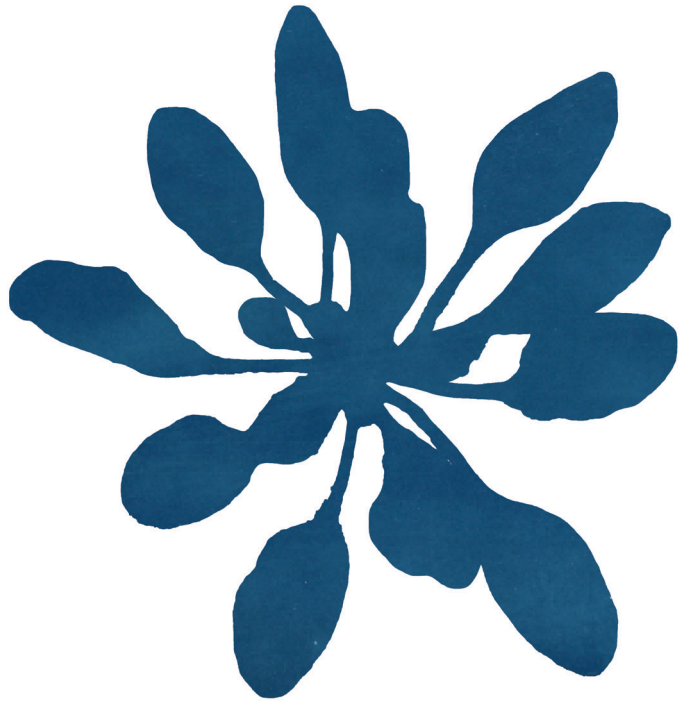
Supplementary figure 2.2: HR-MAS ^{1}H-J-resolved spectra of intact Arabidopsis thaliana leaf zoomed between spectral region (A) 0.9–2.0 ppm and (B) 3.0 and 4.0 ppm showing the assignment of L-isoleucine, L-leucine, L-valine (A) and glucose (B). The J-resolved spectra were measured using pulse sequence ("Jresqfpr") from Bruker's standard pulse program library. Reference spectra of L-isoleucine, L-leucine, L-valine and glucose obtained from have been overlaid for assignment purpose.



Supplementary figure 2.3: ^{1}H-^{1}H COSY spectrum of intact Arabidopsis thaliana leaf zoomed between 3.0 and 5.5 ppm region showing (A) the assignment of glucose, sucrose and fructose and (B) the assignment of α-glucose and β-glucose. Reference spectra have been overlaid for assignment purpose.



Supplementary figure 2.4: Stacked plots of 1D HR-MAS NMR spectra from intact Arabidopsis thaliana leaves. Leaves were collected at different time-points throughout 24h light/dark period showing changes in the levels of metabolites during circadian cycle. Time-points are same as mentioned in Figure 2.1.



Different growth-defence trade-off in *Arabidopsis thaliana* mutants with enhanced growth characteristics

D. Augustijn
N. van Tol
B.J. van der Zaal
H.J.M. de Groot
A. Alia

Submitted to PLoS ONE

3.1 Abstract

Smart crops which yield more biomass to meet the increasing demand of plant biomass has been an active area of research in last few decades. We investigate the metabolic alteration in *Arabidopsis thaliana* VP16-02-003 and VP16-05-014 mutants with a larger rosette surface area phenotype compared to the wild-type Col-0, obtained by genome wide interrogation of transcription patterns. The metabolic profile was obtained directly from intact *Arabidopsis* leaves using high-resolution magic angle spinning (HR-MAS) NMR. Multivariate analysis showed significant alteration of metabolite levels between the mutants and the wild-type Col-0. Interestingly, most of the metabolites that were reduced in VP16-02-003 and VP16-05-014 mutants are generally involved in the defence against stress. These results suggest that diminished defence response leads to an altered growth-defence trade-off in the phenotypically engineered mutants. Our results indicate that growth-defence trade-off is an essential factor for generating crops with improved growth characteristics.

3.2 Introduction

During the last few decades, the demand for agricultural products has increased dramatically.[1-3] In order to meet the actual food demand in 2050, a 70% increase of the food production has to be realized in the coming three decades.[4,5] A possible way to meet this demand is to develop smart crops, varieties which can give more yield with fewer inputs.[5,6] This would also reduce the need for chemicals such as pesticides and fungicides.

In the past, new crop varieties with improved agronomic traits were developed by traditional breeding methods such as hybridization or spontaneous variations.[7] The development can be accelerated by the use of new genome-editing technologies like artificial transcription factors (ATFs), transcription activator-like effector nucleases (TALENs) and CRISPR/CAS9.[8,9] Recently, we have explored genome interrogation using zinc finger artificial transcription factors (ZF-ATFs) as a novel technique to drastically modify genome-wide transcription patterns and to generate novel phenotypes of interest in the model plant species *Arabidopsis*.[9] In these studies, arrays of three zinc fingers (3F) were fused to the transcriptional activation domain of the VP16 protein of the herpes simplex virus.[10] Any 3F motif can recognize 9 base pairs of DNA, corresponding to approximately 1000 recognition sites in the nuclear *Arabidopsis* genome. Expression of a single 3F-VP16 fusion under control of the meristematic RPS5α promoter can thus lead to transcriptional activation of a large number of genomic loci, and consequently to drastic metabolic and phenotypic changes.[9,11] Previously, we have screened a population of transgenic *Arabidopsis* plants harbouring 3F-VP16 encoding gene constructs using enhanced rosette surface area (RSA) as a selection criterion for enhanced overall biomass accumulation. From this phenotypic screen we isolated two novel mutants with significantly larger RSA compared to the wild-type Col-0, and expressing two different 3F-VP16 fusions: VP16-02-003 and VP16-05-014.[12] In that previous study, a transcriptomics analysis was also performed to investigate the changes in the gene expression patterns underlying the enhancement of RSA for both mutants.[12] Interestingly, we observed that, among the differentially expressed genes in both mutants as compared to Col-0, 237 were shared between the mutants, indicating an overlap in the transcriptional mechanism underlying the increase in RSA. In particular, the 93 shared downregulated genes were found to be involved in pathways related to sulfate metabolism

and the further synthesis of the sulfur containing glucosinolates. Additionally, these downregulated genes were also involved in several defence processes, including response to stress (GO:0006950), response to external stimulus (GO:0009605), response to wounding (GO:0009611), response to endogenous stimulus (GO:0009719), response to jasmonic acid (GO:0009753), response to stimulus (GO:0050896), defence response by cell wall thickening (GO:0052482) and defence response by callose deposition in cell wall (GO:0052544). These results suggest that the VP16-02-003 and VP16-05-014 mutants both have reduced ability to deal with stress.[12]

For a comprehensive understanding of newly developed smart plant genotypes, a systems biology approach is indispensable. Using this approach, a plant is seen as a system of interacting units that can be analysed as a whole rather than focusing on individual changes.[6,13] One of the system biology approaches is the metabolomics approach, which aims to determine small molecules that are involved in various physiological functions, such as growth, productivity and defence.[14] Directly examining the metabolic profiles of intact *Arabidopsis* leaves without any extraction is important to understand the functional framework of metabolism in the native state of the leaves. Recently, we have established high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) to obtain the metabolic profile directly from intact wild-type *Arabidopsis* leaves.[15]

In this study, we applied one-dimensional HR-MAS NMR to obtain the metabolic profile directly from the intact leaves of wild-type Columbia (Col-0) *Arabidopsis* plants, and of the VP16-02-003 and VP16-05-014 mutants with enhanced growth characteristics and putatively higher sensitivity to biotic stress based on transcriptomics data. Through metabolic profiling in the native state in combination with multivariate analysis, we here provide novel insights into the biochemical pathways underlying the enhanced rosette surface area phenotype for both mutants.

3.3 Materials & Methods

Plant materials

Arabidopsis thaliana plants were grown in soil and cultivated in a growth chamber maintained at 293 K, 70% relative humidity and at a 12 hours light (200 µmol m^{-2} s^{-1} photosynthetically active radiation) and 12 hours dark regime.[15] Experiments were performed using the *Arabidopsis thaliana* accession Columbia-0 (Col-0) as wild-type. The VP16-02-003 and the VP16-05-014 mutant (both T_3 generation) with an increased rosette surface area were obtained by phenotypic screening of a population of transgenic *Arabidopsis* plants harbouring 3F-VP16 encoding T-DNA constructs, as described previously.[12]

Pigment analysis

To determine the chlorophyll (Chl) a and b and carotenoids concentration, the pigments from the leaves were extracted in DMF (N,N'-dimethylformamide).[16] Absorbance was measured at 750, 663.8, 646.8 and 480 nm using a Shimadzu UV-1700 spectrometer. The $A_{663.8}$, $A_{646.8}$ and A_{750} were used in the equation as described by Porra[17] to calculate the concentrations of Chl *a*, Chl *b* and Chl *a+b*. The carotenoids concentration was determined according to Wellburn.[18]

Quantification of free amino acids, soluble sugars, proteins and starch

The soluble sugar content was determined using the phenol-sulphuric acid method at a wavelength of 490 nm.[19] Glucose ranging from 0 to 250 µg/mL was used to obtain a standard curve. The free amino acids content was determined using the ninhydrin method as described previously.[20] Proteins were extracted from the leaves as described before.[21] Protein content was quantified by a Bradford assay.[22] The starch content of the leaves was measured by determining the glucose released with α-amylase and amyloglucosidase as described by Smith and Zeeman.[23]

Statistical analysis

The data for each independent experiment were subjected to the Student's t-test. The OriginPro 2016 software (Northampton, USA) was used to determine the differences between Col-0 and the VP16-02-003 and VP16-05-014 mutants. Values are presented as means ± standard error (SEM) and statistical significance was determined at $p < 0.05$.

HR-MAS NMR-based metabolic profiling

The leaves were harvested from the plants at 28 days after germination (dpg), frozen immediately in liquid nitrogen and stored at -80 °C until use. A single leaf (0.0684 ± 0.0087 mg, n = 8) was inserted into a 4 mm zirconium oxide (ZrO_2) rotor. 10 µL of deuterated phosphate buffer (100 mM, pH 6) containing 0.1% (w/v) 3-trimetylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) was added as a lock solvent and NMR reference. ^1H high-resolution magic angle spinning (HR-MAS) NMR experiments were performed with a Bruker DMX 400 MHz spectrometer operating at a resonance frequency of 399.427 MHz. The instrument is equipped with a 4 mm HR-MAS dual inverse $^1H/^{13}C$ probe with a magic angle gradient. Data were collected with a spinning frequency of 4 kHz at a temperature of 277 K.

The one-dimensional ^1H HR-MAS NMR spectra were recorded using a rotor synchronized Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence with water suppression.[24] Each one-dimensional spectrum was acquired applying 256 transients, a spectral width of 8000 Hz, a data size of 16 K points, an acquisition time of 2 seconds and a relaxation delay of 2 seconds. The free induction decays (FIDs) were exponentially weighted with a line broadening of 1 Hz. Spectra were phased manually and automatically baseline corrected using TOPSPIN 2.1 (Bruker Analytische Messtechnik, Germany). A gradient-enhanced two-dimensional ^1H-^1H-COSY sequence was applied in order to confirm signal assignments as described before.

Multivariate analysis

A bucket table was generated from the one-dimensional spectra using AMIX software (version 3.8.7, BrukerBioSpin). The region between 4.20 – 6.00 ppm was excluded from the analysis to remove the large water signal. The one-dimensional CPMG spectra were normalized to the total intensity and binned into buckets of 0.04 ppm. The data was mean-centred and the Pareto scaling method was used.[25] Unsupervised PCA and supervised OPLS-DA were performed on the bucket table using the SIMCA software package version 14.0 (Umetrics, Umeå, Sweden). The quality of these models was evaluated by the R^2X and R^2Y, the goodness-of-fit parameters, and Q^2, a measure of the quality of the model based on cross-validation.[26,27] One sample from the VP16-05-014 dataset was removed as it was

a significant outlier defined as an observation located outside the 95% confidence region of the Hotelling's T^2 ellipse in the PCA scatter plot (Supplementary figure 3.1).[28] Further analysis was performed without this outlier. OPLS-DA was used to determine the buckets which are different between the mutants and the wild-type Col-0. In addition, two OPLS-DA models were constructed for each mutant; Col-0 vs VP16-02-003 and Col-0 vs VP16-05-014 (Supplementary figure 3.2). The shared and unique structures between these two OPLS-DA models were investigated using a SUS (shared and unique structures) plot.[29]

Biomarker identification

The metabolites corresponding to the peaks of interest in the determined buckets were identified by the Biological Magnetic Resonance Data Bank (BMRB), Platform for RIKEN Metabolomics[30,31] and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada).

Quantification of metabolites

Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada) was used for quantitative NMR data analysis. The concentrations of the various metabolites in the spectra of intact leaf from wild-type and mutant *Arabidopsis* were determined by the known concentration of the reference peak of TSP. All Student's t-test analyses of the NMR quantification results were performed with OriginPro 2016 (Northampton, USA).

3.4 Results and Discussion

Pigment characterisation of the Arabidopsis mutants

Pigments are involved in the absorption of light, electron transfer and other important processes in photosynthesis.[16] Photosynthetic pigment content of leaves of Col-0 and *Arabidopsis* VP16-02-003 and VP16-05-014 plants has been analysed (Table 3.1). As shown in Table 3.1, the VP16-02-003 mutant shows a significantly higher level of chlorophyll *b* as compared to Col-0. This results in a lower chlorophyll *a/b* ratio of VP16-02-003 mutant (1.85 ± 0.12) in comparison to Col-0 (3.55 ± 0.23, $p < 0.05$). Interestingly, the chlorophyll *b* concentration was 67.2% higher in VP16-05-014 as compared to Col-0 ($p < 0.05$). The chlorophyll *a/b* ratio was thus significantly lower in VP16-05-014 mutant as compared to Col-0 ($p < 0.05$).

*Table 3.1: Pigment contraction of Arabidopsis thaliana Col-0, VP16-02-003 and VP16-05-014 in µg/mL g^{-1} fresh weight. Data is expressed as mean ± SEM. * $p < 0.05$ compared with Col-0.*

Concentration (µg/mL g^{-1} fresh weight)	Col-0	VP16-02-003	VP16-05-014
Chlorophyll *a*	451.70 ± 55.13	354.32 ± 46.76	462.34 ± 68.78
Chlorophyll *b*	126.43 ± 13.76	196.20 ± 29.67 *	211.42 ± 26.10 *
Chlorophyll *a+b*	578.13 ± 67.62	550.52 ± 75.00	673.76 ± 94.39
Chlorophyll *a/b*	3.55 ± 0.23	1.85 ± 0.12 *	2.17 ± 0.07 *
Carotenoids	76.80 ± 11.92	80.70 ± 10.11	108.74 ± 10.36

The reduced chlorophyll *a/b* ratio in both mutants may indicate an increase in grana stacking as has been proposed in earlier studies.[32,33] For both mutants, the decreased ratio of chlorophyll *a/b* can be attributed to an increased chlorophyll b content since the chlorophyll a concentration is not changed significantly in both mutants. Increased chlorophyll *b* content is an indication of a larger core light-harvesting complex.[33,34] A low chlorophyll *a/b* ratio has been reported in *Arabidopsis* under low-light conditions in previous studies.[35] The *Arabidopsis thaliana* BC mutant accumulates higher concentrations of chlorophyll *b* and the total chlorophyll content in this mutant was much lower than for the Col-0. The deregulation of chlorophyll *b* in the BC mutant impairs its ability to deal with different light conditions.[34]

Levels of free amino acids, soluble sugars, proteins and starch

Prior to metabolic profiling, the concentration of free amino acids, proteins, soluble sugars and starch was determined for extracts of leaves from the VP16-02-003 and VP16-05-014 mutant and Col-0. Interestingly, an overall decline in free amino acids, soluble sugars, proteins, as well as starch content, was observed for the VP16-02-003 and VP16-05-014 mutants of *Arabidopsis* as compared to Col-0 (Table 3.2).

*Table 3.2: Total free amino acids, protein, soluble sugar and starch content in mg/g fresh weight for leaves of Arabidopsis thaliana Col-0, VP16-02-003 and VP16-05-014. Data is expressed as mean ± SEM. * $p < 0.05$ compared with Col-0.*

Content (mg/g FW)	Col-0	VP16-02-003	VP16-05-014
Free amino acids	1.48 ± 0.05	1.17 ± 0.05 *	1.04 ± 0.01 *
Protein	1.57 ± 0.06	1.34 ± 0.08 *	0.98 ± 0.01 *
Soluble sugar	0.18 ± 0.01	0.11 ± 0.01 *	0.13 ± 0.01 *
Starch	0.57 ± 0.02	0.23 ± 0.01 *	0.31 ± 0.02 *

Sulpice *et al.* observed a negative correlation between biomass and the levels of starch, total protein and total free amino acids in *Arabidopsis thaliana*.[36] The level of soluble sugars such as sucrose has also found to be negatively correlated with biomass.[36] Sugars, such as glucose and sucrose are important products of photosynthesis and play an essential role in controlling plant growth, development and defence.[14] The decline in the level of soluble sugars, accompanied by an increase in overall biomass of the two mutants indicates that sugar resources are likely diverted toward growth and storage products in these mutants, rather than defence related processes.

Metabolic profiling of the VP16-02-003 and VP16-05-014 mutant

In order to identify metabolites and biochemical pathways responsible for the increased rosette surface area shared phenotype of the VP16-02-003 and VP16-05-014 mutant, their metabolic profiles have been analysed for intact leaves using HR-MAS NMR. In Figure 3.1, representative one-dimensional ^{1}H NMR spectra of Col-0, VP16-02-003 and VP16-05-014 are shown. Two-dimensional ^{1}H-^{1}H COSY enabled confirmation of metabolites by their spin systems. The signals of various metabolites were assigned with the help of literature data from the Biological Magnetic Resonance Data Bank (BMRB).[37,38]

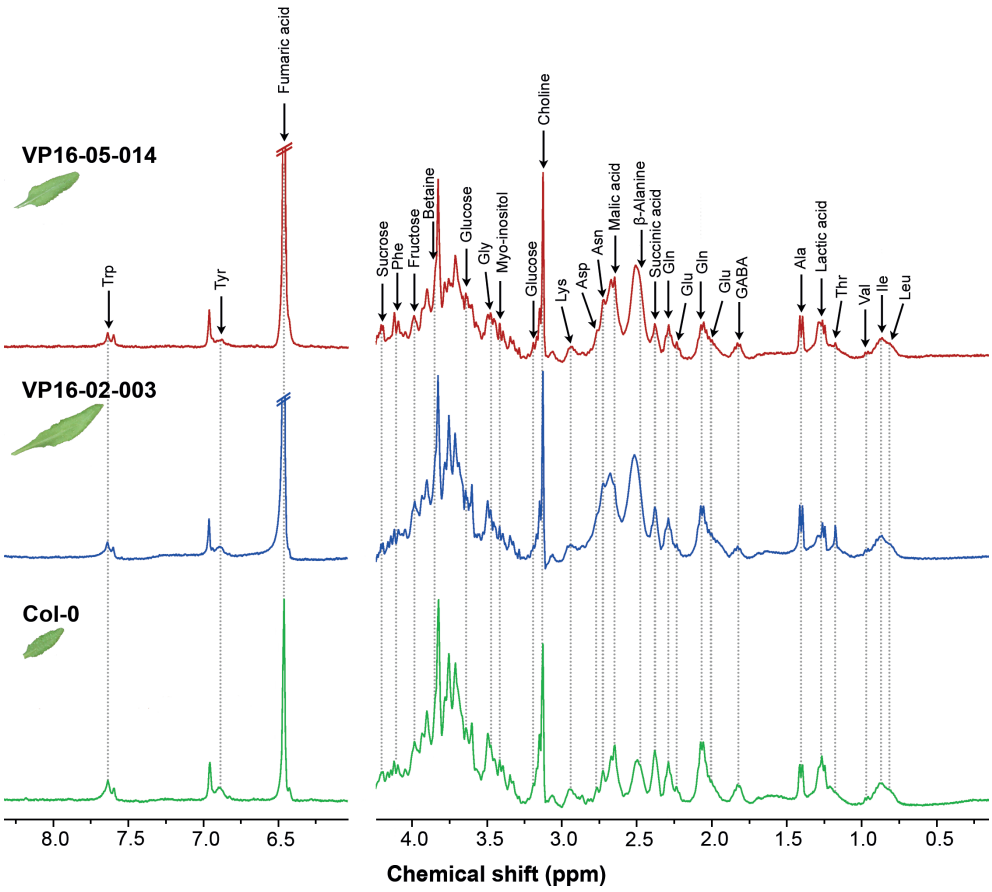


Figure 3.1: Representative regions of one-dimensional 400 MHz ^{1}H CPMG NMR data collected from intact leaves of Arabidopsis thaliana Col-0 (bottom panels), VP16-02-003 (middle panels) and VP16-05-014 (top panels) grown in a 12h-light/12h-dark regime obtained by ^{1}H HR-MAS NMR. The intact leaves were harvested 28 dpg at t = 6 hours (6 hours after the beginning of the light period). The signals from the metabolites with the highest concentrations, have been assigned as indicated in the spectra.

Identification of the increased rosette surface area shared phenotype

To probe if *Arabidopsis thaliana* Col-0, VP16-02-003 and VP16-05-014 can be discriminated from each other based on their metabolic profiles, multivariate analysis was applied to the HR-MAS spectra from both mutants and Col-0. Unsupervised PCA was performed which explained 76.1% of the variation by a three-component model (see Figure 3.2A), which shows a clear group separation of VP16-05-014 from the Col-0 and VP16-02-003. In contrast, there was no clear group separation between Col-0 and VP16-02-003 in the PCA score plot. Supervised OPLS-DA was applied to further understand the separation between the wild-type Col-0, VP16-02-003 and VP16-05-014 and to identify crucial biomarker candidates involved in the increased rosette surface area phenotype. Figure 3.2B shows the score plot

of the OPLS-DA. The R^2X, R^2Y and Q^2 were 0.849, 0.942 and 0.603, respectively. The OPLS-DA model was found to be of good quality and has an accurate prediction. The score plot also shows that the biological variation for the VP16-02-003 or VP16-05-014 is less than for the wild-type *Arabidopsis* Col-0.

A Shared and Unique Structures (SUS) plot can be a powerful method to identify potential biomarkers for the enhanced growth characteristics for both mutants.[29] To obtain a SUS plot, two separate OPLS-DA models (Col-0 versus VP16-02-003 and Col-0 versus VP16-05-014) were generated from the metabolic profiles (Supplementary figure 3.2). The correlation coefficients of the predictive component p(corr) for both models are plotted against each other in Figure 3.2C. Concentrations of metabolites plotted in the upper right corner of the SUS-plot increased and those plotted in the lower left corner decreased in both mutants as compared to Col-0. The upper left corner and lower right corner of the SUS-plot contain metabolites with anti-correlated concentrations in the two mutants in comparison to Col-0. Eighteen biomarkers were determined from the SUS plot which show variation between Col-0, VP16-02-003 and VP16-05-014 mutants, including organic acids (fumaric acid, malic acid, lactic acid), sugars (fructose, glucose), a sugar alcohol (myo-inositol), precursor of cell wall components (choline), an organic osmolyte (betaine) and free amino acids (L-alanine, β-alanine, L-asparagine, L-aspartic acid, L-glutamic acid, L-glutamine, L-glycine, L-lysine, L-phenylalanine and L-tyrosine). Most of the identified biomarkers are primary metabolites directly involved in essential processes as growth, development and defence.[39,40] This provides evidence that the larger rosette surface area phenotype of both mutants involves re-allocation of primary resources.

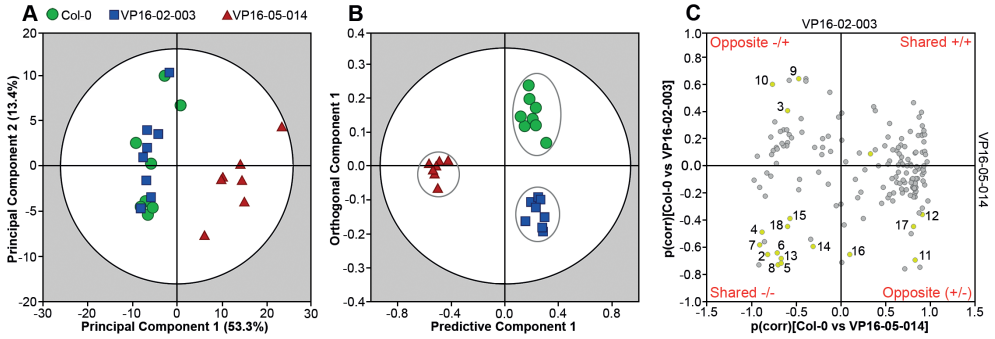


Figure 3.2: Multivariate analysis of ^{1}H HR-MAS NMR metabolic data collected from Arabidopsis thaliana Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲). (A) Principal component analysis (PCA) score plot with R^2X = 0.761 and Q^2 = 0.611. The dark circle shows the 95% confidence interval using Hotelling T^2 statistics. (B) Orthogonal partial least squares discriminant analysis (OPLS-DA) score plot with R^2X = 0.849, R^2Y = 0.942, Q^2 = 0.603, which indicates separation between the Col-0, VP16-02-003 and VP16-05-014 Arabidopsis plants based on their metabolic profile. The dark circle represents the Hotelling T^2 interval with 95% confidence. (C) SUS plot, corresponding to the OPLS-DA model, represents biomarkers responsible for the separation in the score plot. 1. Fumaric acid; 2. Malic acid; 3. Lactic acid; 4. Fructose; 5. Glucose; 6. Myo-inositol; 7. Choline; 8. Betaine; 9. L-alanine; 10. β-alanine; 11. L-asparagine; 12. L-aspartic acid; 13. L-glutamic acid; 14. L-glutamine; 15. L-glycine; 16. L-lysine; 17. L-phenylalanine; 18. L-tyrosine.

Metabolic evidence for altered growth-defence trade-off in the VP16-02-003 and VP16-05-014 mutants

The quantitative analysis of the metabolites that show significant variation between Col-0, VP16-02-003 and VP16-05-014 plants is shown in Figure 3.3.

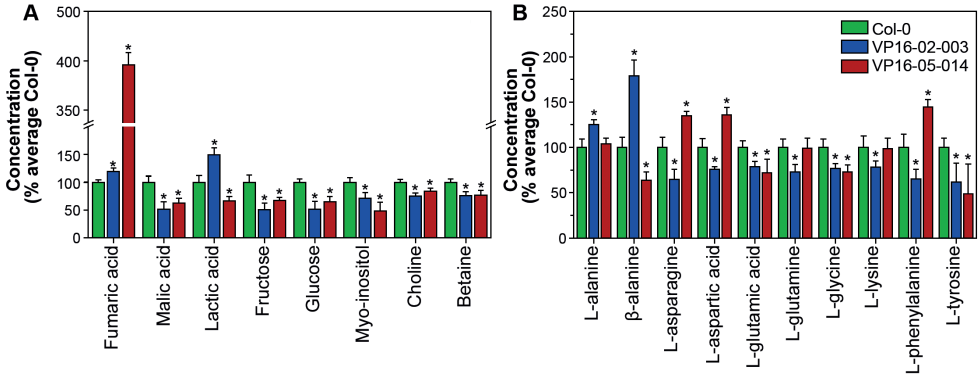


Figure 3.3: Metabolic alterations in the leaf of Arabidopsis thaliana Col-0, VP16-02-003 and VP16-05-014 plants grown in a 12h-light/12h-dark regime by ^{1}H HR-MAS NMR. (A) Relative levels of organic acids, sugars, sugar alcohol, precursor of cell wall components and organic osmolyte in leaves of VP16-02-003 and VP16-05-014 plants in comparison to the level in Col-0. (B) Relative levels of free amino acids in leaves of VP16-02-003 and VP16-05-014 plants in comparison to Col-0. Concentrations are represented by the mean ± SEM averaged over n = 8 samples. The asterisks (*) indicate significant differences between concentrations for Col-0 and mutants, calculated with Student's t-test ($p < 0.05$).

Organic acids, like fumaric acid and malic acid, play an important role in the major carbon metabolism involving glycolysis, the tricarboxylic acid (TCA) cycle, and the photorespiration cycle.[41,42] The fumaric acid level was significantly elevated by 19.5% in the VP16-02-003 mutant and by 295.8% in the VP16-05-014 mutant (Figure 3.3A). The primary metabolite fumaric acid participates in multiple pathways in plant metabolism and is considered to be one of the major forms of fixed carbon in some C3 plants, including *Arabidopsis*.[15,43] In particular, fumaric acid can accumulate to levels of several milligrams per gram fresh weight in *Arabidopsis* leaves, often exceeding concentrations of starch and soluble sugars.[43]

In contrast to fumaric acid, the malic acid level is reduced for both mutants (Figure 3.3A). Malic acid is involved in various physiological functions in the plant cells, such as supplying NADH for nitrate reduction and carbon skeletons, and delivering NADPH for fatty acid biosynthesis.[44] The observed reduced malic acid level for both mutants can be a consequence of its enhanced utilization in downstream pathways involved in the growth promoting phenotype.[45] The role of lactic acid in the leaves of *Arabidopsis thaliana* is not very clear. It has been reported to play a role in plant defence against pathogens.[45] Also a growth promoting effect of lactic acid has been reported earlier.[46,47] The lactic acid concentration was elevated by 49.1% in the VP16-02-003 mutant ($p < 0.05$), and reduced by 33.2% in the VP16-05-014 mutant ($p < 0.05$) (Figure 3.3A). Hence the higher level of lactic acid in the VP16-02-003 mutant can be involved in establishing the growth promoting phenotype of this mutant.

Primary sugars in plants such as glucose, fructose, and sucrose, are produced during photosynthesis, provide the primary energy supply and serve as storage metabolites in plants. These sugars have a regulatory role in photosynthesis, growth and development and as a signalling molecule to modulate gene expression.[48] The levels of fructose and glucose were in both the VP16-02-003 and the VP16-05-014 mutants significantly decreased in comparison to Col-0 (Figure 3.3A). Although the stress response is a very dynamic process and differs for every stress type, soluble sugar concentrations are common denominators of defence and strongly decrease in response to different forms of abiotic stress, as energy is needed to operate defence mechanisms. In general low sugar concentrations lead to an impaired stress response.[48,49]

Myo-inositol is a signalling metabolite in *Arabidopsis thaliana*.[50] It is involved in stress response, regulation of cell death and cell wall biosynthesis.[51,52] The concentration of myo-inositol was reduced by 28.3% ($p < 0.05$) in the VP16-02-003 mutant and by 51.7% in the VP16-05-014 mutant ($p < 0.05$) (Figure 3.3A). A reduced pool of myo-inositol in the VP16-02-003 and VP16-05-014 mutant may reflect an impaired defence pathway regulation in these mutants. For instance, in a previous study, the reduced level of myo-inositol was observed in the *mips1* mutant, a mutant which shows increased sensitivity to reactive oxygen species stress.[51]

Choline is an important precursor for membrane phospholipids in plants. Choline can be oxidized in a 2-step reaction via betaine aldehyde to betaine (glycine betaine, N,N,N-trimethylglycine). Accumulation of betaine in *Arabidopsis thaliana* leads to a higher tolerance for abiotic stress.[53,54] The choline and betaine levels in the VP16-02-003 and the VP16-05-014 mutant were both significantly reduced in comparison to the wild-type Col-0 (Figure 3.3A). This is in line with less available resources for stress resistance and defence and low concentrations of the soluble primary carriers malic acid, fructose and glucose, as well as more storage in the form of fumaric acid.

Amino acids are essential precursors for a wide range of cellular components like proteins, nucleotides, chlorophylls and nitrogen-containing compounds.[55,56] Figure 3.3B shows the concentrations of free amino acids in the VP16-02-003 and the VP16-05-014 mutant in comparison to *Arabidopsis* Col-0. The pattern of free amino acid levels differs between the two mutants. The levels of L-alanine and β-alanine were significantly higher in the VP16-02-003 mutant in comparison to the wild-type Col-0. An increase in alanine under sulphur limitation has been reported earlier.[57] Transcriptome analysis of this mutant has revealed downregulation of sulphate metabolism genes.[12] High levels of alanine in the VP16-02-003 mutant may thus be connected to reduced sulphur fixation. The concentrations of free amino acids such as L-asparagine, L-aspartic acid, L-glutamic acid, L-glutamine, L-glycine, L-lysine, L-phenylalanine and L-tyrosine were significantly lower in VP16-02-003 in contrast to Col-0. In the VP16-05-014 mutant, levels of L-asparagine, L-aspartic acid and L-phenylalanine were increased while the concentrations of β-alanine, L-glutamic acid, L-glycine and L-tyrosine were decreased relative to Col-0. From earlier studies, it is known that a decreased level of sugar leads to the inhibition of amino acid biosynthesis. The patterns we observe for both mutants are in line with this scenario.[55] In addition, amino acid metabolism plays a regulatory role in the response to stress.[56,58] In conclusion, a decline in the levels of free

amino acids for both mutants reflects their reduced investments to defence responses against stress.

The changes of metabolic profiles for the VP16-02-003 and the VP16-05-014 mutant that share a larger rosette surface area phenotype are mostly associated with the response to stress. These results are in line with our earlier transcriptome analysis, which shows that many genes that are downregulated in both mutants are involved in defence processes.[12] Adaptation of the physiological function in *Arabidopsis thaliana* requires balancing of primary metabolites.[59] The trade-off between plant growth and defence implies that both are negatively correlated.[59-61] Hence the growth will improve when the defence system is less active in plants. Our results show that the VP16-02-003 and the VP16-05-014 mutants have a reduced defence response against stress (Figure 3.4), which shifts the balance toward enhanced growth.

The fundamental understanding about linking the gene regulatory network to the physiological response obtained in this work paves the way for further investigations. In particular it is known that physiological functions in *Arabidopsis* are regulated by the circadian cycle[62,63] and analysis of rhythmic patterns of the biomarkers in the VP16-02-003 and the VP16-05-014 mutants may help to resolve the underlying mechanisms involved in growth-defence trade-off.

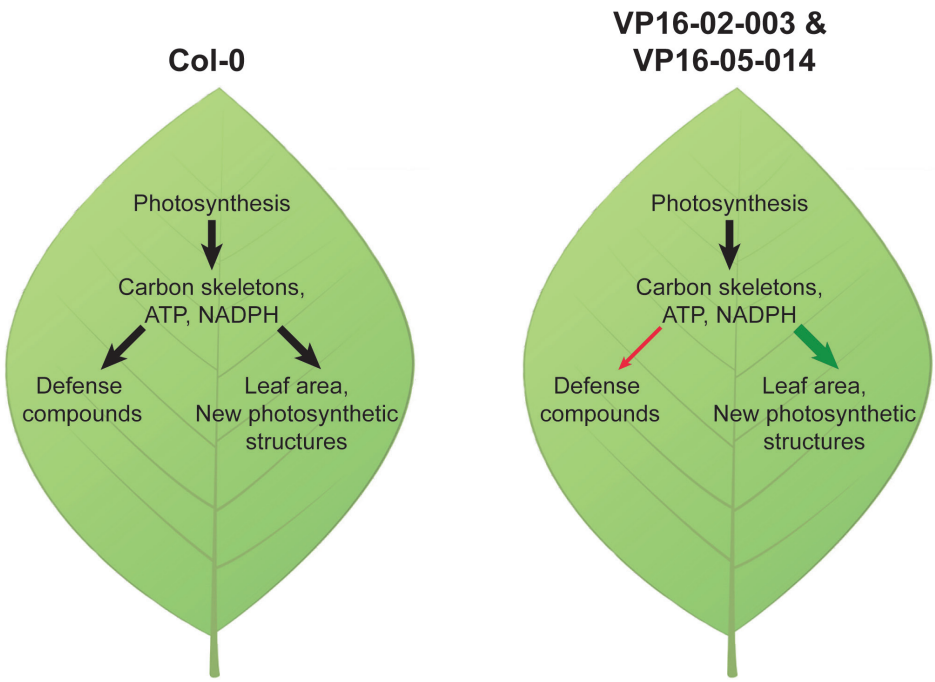


Figure 3.4: Schematic overview of the different growth-defence trade-off in the VP16-02-003 and VP16-05-014 mutant with a larger rosette surface area shared phenotype in comparison to wild-type Arabidopsis thaliana Col-0.

3.5 Conclusion

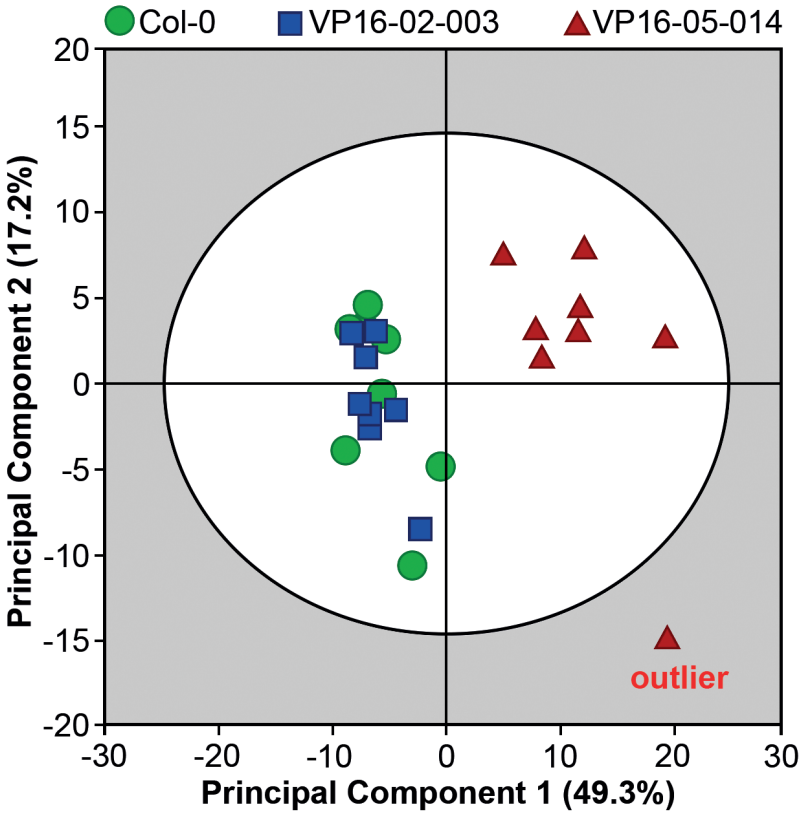
In this study, the metabolic profiles of the *Arabidopsis thaliana* VP16-02-003 and the VP16-05-014 mutants with a larger rosette surface shared phenotype are investigated with the HR-MAS NMR-based metabolomics approach. The results provide converging evidence that the alteration of the metabolic profile of both mutants is due to lower defence responses against stress. Growth-defence trade-off appears to be an essential factor in generating crop with improved growth characteristics.

3.6 References

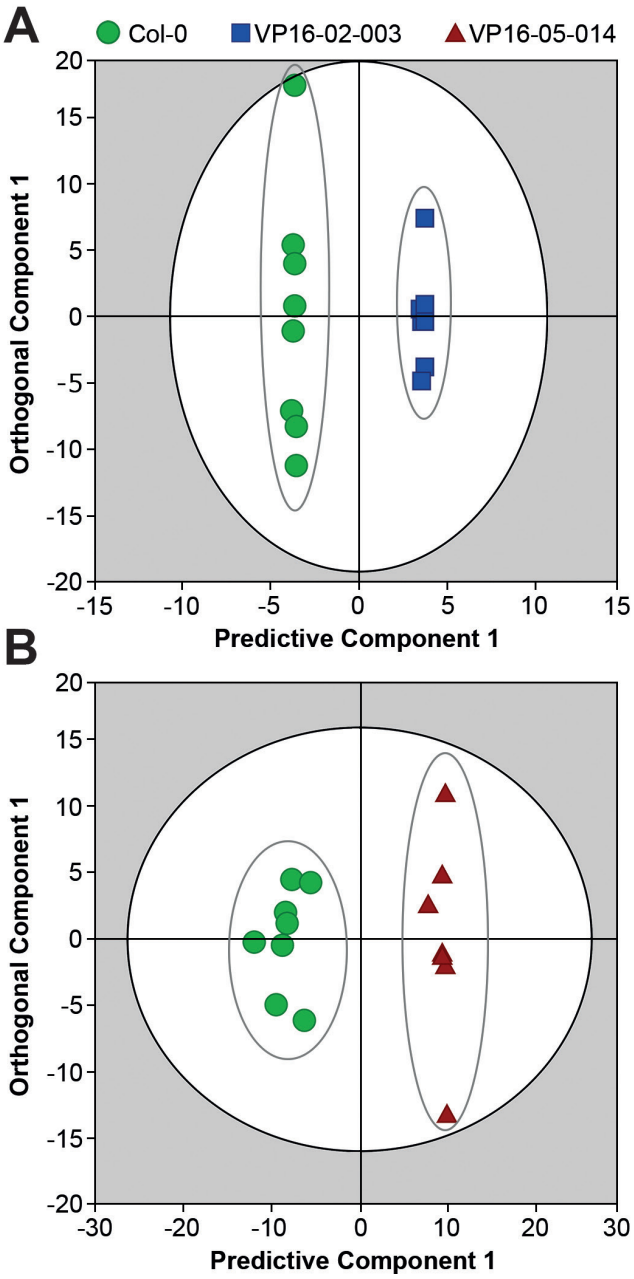
[1] M. D. Edgerton, Plant Physiology 2009, 149, 7–13.
[2] H. Haberl et al., Current Opinion in Environmental Sustainability 2010, 2, 394–403.
[3] D. Tilman et al., Nature 2002, 418, 671–677.
[4] K. F. Davis et al., Global Environmental Change 2016, 39, 125–132.
[5] C. Mba et al., Agriculture & Food Security 2012, 1, 7–17.
[6] A. Kumar et al., OMICS 2015, 19, 581–601.
[7] C. M. Rommens et al., Trends in Plant Science 2007, 12, 397–403.
[8] G. Andolfo et al., Front. Plant Sci. 2016, 7, 19035.
[9] N. e Tol, B. J. van der Zaal, Plant Science 2014, 225, 58–67.
[10] J. Li et al., Plant Biotechnology Journal 2013, 11, 671–680.
[11] B. I. Lindhout et al., The Plant Journal 2006, 48, 475–483.
[12] N. van Tol et al., PLoS One 2017, 12, e0174236.
[13] B. P. Sheth, V. S. Thaker, Planta 2014, 240, 33–54.
[14] R. C. Meyer et al., PNAS 2007, 104, 4759–4764.
[15] D. Augustijn et al., PLoS One 2016, 11, e0163258.
[16] X. Hu et al., Plant Methods 2013, 9, 1–1.
[17] R. J. Porra, Photosynth Res 2002, 73, 149–159.
[18] A. R. Wellburn, Journal of Plant Physiology 1994, 144, 307–313.
[19] M. DuBois et al., Anal. Chem. 1956, 28, 350–356.
[20] S. Gepstein, K. V. Thimann, Plant Physiology 1981, 68, 349–354.
[21] Y. Gibon et al., Plant Cell 2004, 16, 3304–3325.
[22] M. M. Bradford, Analytical Biochemistry 1976, 72, 248–254.
[23] A. M. Smith, S. C. Zeeman, Nat. Protocols 2006, 1, 1342–1345.
[24] S. Meiboom, D. Gill, Review of Scientific Instruments 1958, 29, 688–691.
[25] L. R. Euceda et al., Scand J Clin Lab Invest 2015, 75, 193–203.
[26] B. Worley, R. Powers, CMB 2012, 1, 92–107.
[27] J. A. Westerhuis et al., Metabolomics 2008, 4, 81–89.
[28] J. Trygg et al., J. Proteome Res. 2007, 6, 469–479.
[29] S. Wiklund et al., Anal. Chem. 2008, 80, 115–122.
[30] K. Akiyama et al., In Silico Biol 2008, 8, 339–345.
[31] T. Sakurai et al., Plant Cell Physiol 2013, 54, e5–e5.
[32] R. E. Häusler et al., Plant Physiology 2009, 149, 515–533.
[33] A. K. Biswal et al., Plant Physiology 2012, 159, 433–449.
[34] Y. Sakuraba et al., Plant Cell Physiol 2010, 51, 1055–1065.
[35] R. Kouřil et al., Biochimica et Biophysica Acta (BBA) - Bioenergetics 2013, 1827, 411–419.
[36] R. Sulpice et al., PNAS 2009, 106, 10348–10353.
[37] V. Govindaraju et al., NMR Biomed 2000, 13, 129–153.
[38] E. L. Ulrich et al., Nucleic Acids Res 2008, 36, D402–8.

[39] J. Schwachtje et al., Sci. Rep. 2018, 8, 216.
[40] C. M. Rojas et al., Front. Plant Sci. 2014, 5, 17.
[41] M. B. Zell et al., Plant Physiology 2010, 152, 1251–1262.
[42] A. U. Igamberdiev, A. T. Eprintsev, Front. Plant Sci. 2016, 7, 513.
[43] D. W. Chia et al., Planta 2000, 211, 743–751.
[44] H. Fahnenstich et al., Plant Physiology 2007, 145, 640–652.
[45] V. G. Maurino, M. K. M. Engqvist, Arabidopsis Book 2015, 13, e0182.
[46] A. M. Kinnersley et al., Plant Growth Regulation 1990, 9, 137–146.
[47] M. Yoshikawa et al., Can. J. Microbiol. 1993, 39, 1150–1154.
[48] M. Rosa et al., Plant Signal Behav 2009, 4, 388–393.
[49] I. Morkunas, L. Ratajczak, Acta Physiol Plant 2014, 36, 1607–1619.
[50] F. A. Loewus, P. P. N. Murthy, Plant Science 2000, 150, 1–19.
[51] J. L. Donahue et al., Plant Cell 2010, 22, 888–903.
[52] P. H. Meng et al., PLoS One 2009, 4, e7364.
[53] A. Sakamoto, N. Murata, Plant Cell Environ 2002, 25, 163–171.
[54] J. Huang et al., Plant Physiol Biochem 2008, 46, 647–654.
[55] M. Stitt et al., Journal of Experimental Botany 2002, 53, 959–970.
[56] T. M. Hildebrandt et al., Molecular Plant 2015, 8, 1563–1579.
[57] M. Giordano et al., Plant Physiology 2000, 124, 857–864.
[58] R. Pratelli, G. Pilot, Journal of Experimental Botany 2014, 65, 5535–5556.
[59] S. Caretto et al., Int. J. Mol. Sci. 2015, 16, 26378–26394.
[60] B. Huot et al., Molecular Plant 2014, 7, 1267–1287.
[61] T. Zust, A. A. Agrawal, Annu Rev Plant Biol 2017, 68, 513–534.
[62] C. Bendix et al., Molecular Plant 2015, 8, 1135–1152.
[63] K. Greenham, C. R. McClung, Nat Rev Genet n.d., 16, 598–610.

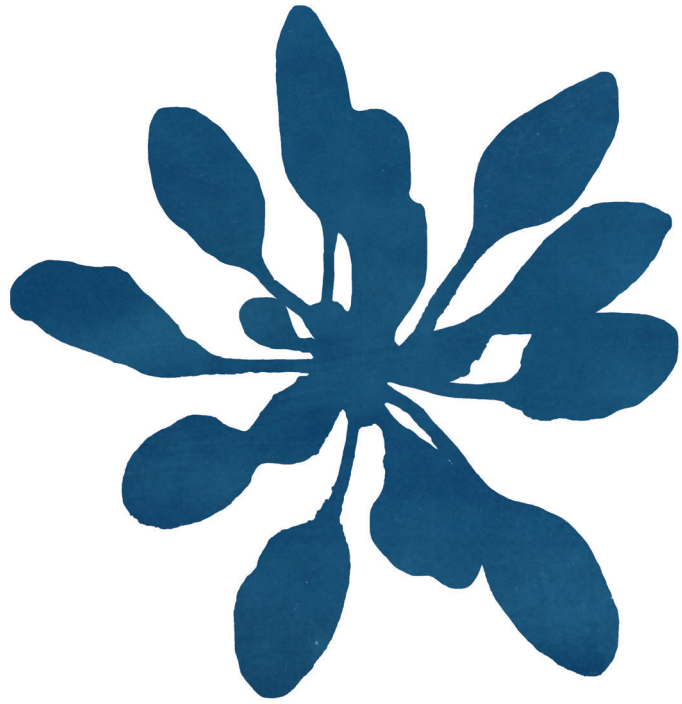
3.7 Supplementary information



Supplementary figure 3.1: Principal component analysis (PCA) score plots derived from one-dimensional ^{1}H HR-MAS NMR spectra of Arabidopsis thaliana Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲) with the outlier for VP16-05-014. A three components model with $R^2X = 0.756$, $Q^2 = 0.585$. The dark ellipse represents the Hotelling T^2 interval with 95% confidence.



Supplementary figure 3.2: Score plot of the orthogonal partial least square-discriminant (OPLS-DA) model derived from ^{1}H HR-MAS spectra of Arabidopsis thaliana Col-0 and VP16-02-003 (A) and Col-0 and VP16-05-014 (B). For model A: $R^2X = 0.865$, $R^2Y = 0.999$, $Q^2 = 0.568$. For model B: $R^2X = 0.666$, $R^2Y = 0.975$, $Q^2 = 0.933$. The dark ellipse shows the 95% confidence interval using Hotelling T^2 statistics.



Contrasting metabolite levels and a robust circadian rhythm in *Arabidopsis thaliana* mutants with enhanced growth characteristics

D. Augustijn
N. van Tol
B.J. van der Zaal
H.J.M. de Groot
A. Alia

In preparation

4.1 Abstract

Climate change and the rising food demand provide a need for smart crops that yield more biomass. A genome-wide interrogation approach was used to obtain the *Arabidopsis thaliana* VP16-02-003 and the VP16-05-014 mutants with enhanced growth characteristics by genome-wide reprogramming of gene expression. The photosynthetic machinery is under control of the circadian clock that can perform anticipative switching of the conversion system to prepare plants during the light/dark cycle. We study the metabolic profile at seven different time-points throughout the circadian cycle using high-resolution magic angle spinning (HR-MAS) NMR to investigate the influences of the circadian clock on the metabolic rhythm of the eighteen biomarkers for the larger rosette surface area phenotype. We found that the circadian rhythm of biomarker metabolites is remarkably robust across wild-type Col-0 and VP16-02-003 and the VP16-05-014 mutants, with widely different metabolite levels of both mutants compared to Col-0 throughout the whole circadian cycle. Our analysis reveals that robustness is achieved through functional independence between the circadian clock and primary metabolic processes.

4.2 Introduction

Smart crops with high biomass yield and with a reduced need for fertilizer and pesticides can help to meet the increasing demand for agricultural products.[1,2] Smart crops can be developed by genome editing tools including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats/Cas9 (CRISPR/Cas9).[3,4] In this study, zinc finger artificial transcription factors (ZF-ATFs) were used to obtain *Arabidopsis thaliana* mutants with enhanced growth characteristics. The transcriptional activator protein VP16 from the herpes simplex virus is fused to an array of three zinc fingers to provide an artificial gene construct, denoted 3F-VP16. Arabidopsis plants are transformed with the 3F-VP16 construct under the control of the promotor of the RPS5α gene. The zinc-fingers can have ±1000 binding sites, both in the coding and in the non-coding DNA of the *Arabidopsis* genome of 130 Mbp. This can lead to drastic changes in genome-wide expression patterns and provide access to rare phenotypes of plants.[4,5] In our previous study, two *Arabidopsis thaliana* mutants, VP16-02-003 and VP16-05-014, with enhanced growth characteristics were obtained using genome interrogation with ZF-ATFs.[6] We have introduced a systems biology approach to resolve primary processes of metabolic regulation and conversion leading to enhanced growth from the complicated biological background.[7]

In the previous work, the metabolic profile for both mutants was studied at one time-point in the middle of the light period of 12 hours in a 24-hour light/dark regime (Chapter 3). The changes in the metabolomics and transcriptomics of both mutants compared to *Arabidopsis* Col-0 gave us insight into the reduced ability to respond to environmental stress of both mutants.[6] Our earlier studies provide converging evidence that the improved growth characteristics of the mutants are a consequence of the reduce defence response in the context of the growth-defence trade-off.[8-10]

It is known that physiological functions, like growth, flowering time, response to stress and metabolism are regulated by the circadian cycle.[11,12] A circadian cycle is a biological

process that displays an endogenous, entrainable oscillation of about 24 hours. Hence the circadian clock comprises three physical entities that fulfil the functional requirements of pre-dawn activation, evening deactivation and the afternoon switching between the morning and evening regimes, providing a nice example of a functional based framework with limited complexity (Figure 4.1A).[13] The molecular mechanism of the circadian clock in *Arabidopsis thaliana* is studied extensively during the last decades and consists of three interlocked transcriptional-translation feedback loops (Figure 4.1B).[11,12,14-16] The central loop of the circadian clock consists of the morning-expressed genes CCA1 (CIRCADIAN CLOCK ASSOCIATED 1) and LHY (LATE ELONGATED HYPOCOTYL) and the evening-expressed gene TOC1 (TIME OF CAB EXPRESSION 1). CCA1 and LHY form a heterodimer which represses the expression of TOC1. The heterodimer also positively regulates the afternoon-expressed pseudo-response regulators PRR5, PRR7 and PRR9. In turn, PRR5, PRR7 and PRR9 repress the transcription of CCA1 and LHY and allows the induction of the evening-expressed gene TOC1. The evening-expressed genes ELF3 (EARLY FLOWERING 3), ELF4 and LUX (LUX ARRYTHMO) will be expressed. These three proteins form the evening complex (EC). The evening complex represses the expression of the PRRs which allows the expression of CCA1 and LHY in the early morning.[11,12,14-16]

According to recent insights into the biological design from a functional perspective, robustness requires functional independence of the abundant subsystems, which, in the case of the clock, are controlled through interfaces while leaving the clock itself undisturbed. I will show this chapter that this is in fact realized in the *Arabidopsis* VP16-02-003 and VP16-05-014 mutants, where gross changes in concentration of primary metabolites are realized while the clock is running the same way across wild-type and mutants.

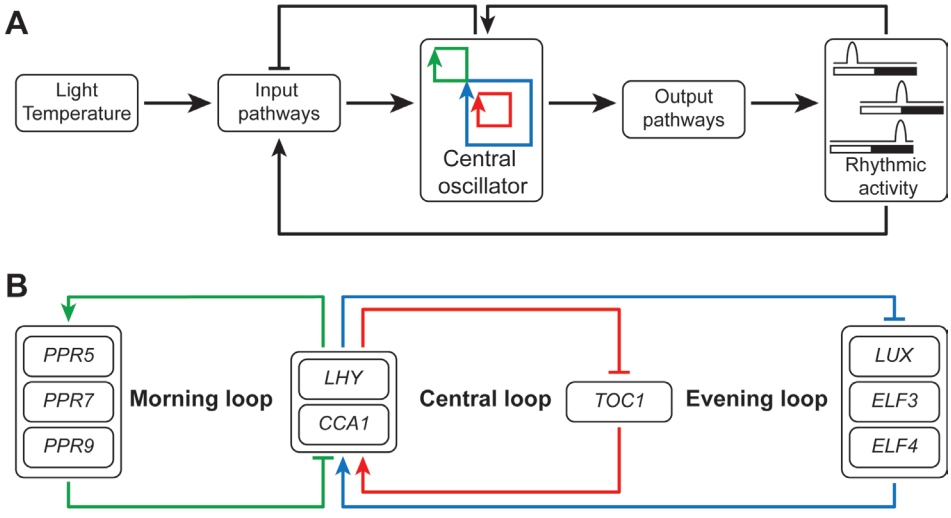


Figure 4.1: Schematic overview of the Arabidopsis circadian clock. A) The central oscillator consists of three interlocked feedback loops. External signals like light and temperature can influence the input pathways for the central oscillator. On his turn, the central oscillator activates output pathways depending on the time of the day. B) Model of the central oscillator containing three interlocked transcriptional-translation feedback loops.

We have studied both the circadian rhythm and the concentration of the metabolites using high-resolution magic angle spinning (HR-MAS) NMR to obtain the metabolic profiles at different time-points of the light/dark cycle for the *Arabidopsis thaliana* wild-type Columbia 0 (Col-0) and two mutants, VP16-02-003 and VP16-05-014. The data provide converging evidence that the clock functional periodicity is independent of mitigation of cellular complexity and growth-defence trade-off.

4.3 Materials & Methods

Plant materials and sample collection

Arabidopsis thaliana plants were grown in soil and cultivated in a climate-controlled growth chamber at 293 K, 70% relative humidity and with a light regime of 12 hours light (200 µmol m^{-2} s^{-1}) and 12 hours dark.[7] The VP16-02-003 and VP16-05-014 mutant with an increased rosette surface area phenotype were investigated in this study in comparison to the wild-type *Arabidopsis* accession Columbia-0 (Col-0).[6] The leaves were harvested at 35 days post germination at seven different time-points throughout the light/dark cycle (Figure 4.2). Samples were frozen immediately in liquid nitrogen and stored at -80˚C.

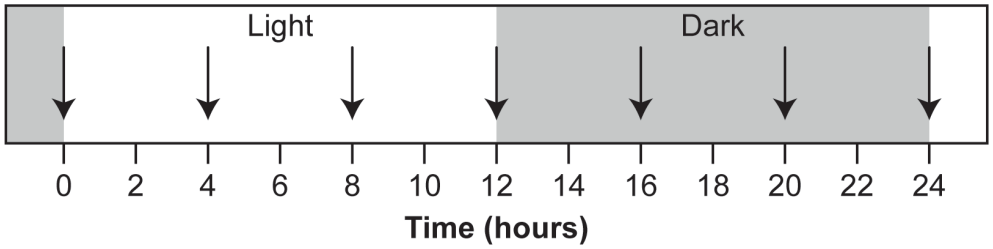


Figure 4.2: Time-points of harvesting of Arabidopsis thaliana Col-0, VP16-02-003 and VP16-05-014 rosette leaves throughout the light/dark cycle.

Quantification of free amino acids, proteins, soluble sugars and starch

The free amino acid content was assayed using the ninhydrin method[17] and the protein content was determined using a Bradford assay.[18] The concentration of soluble sugars was determined using the phenol-sulphuric acid method.[19] The method of Smith and Zeeman was used to determine the starch content of the leaves.[20]

Statistical analysis

The Student's t-test function of OriginPro 2016 software (Northampton, USA) was used to determine the significant differences between Arabidopsis wild-type Col-0, VP16-02-003 and VP16-05-014 for the data for each experiment. Values are presented as means ± standard error (SEM) and statistical significance were determined at $p < 0.05$. The same colours for each class are used throughout the whole chapter.

HR-MAS NMR-based metabolic profiling

A single leaf (n = 6) was inserted into a 4 mm ZrO_2 rotor with 10 µL of deuterated phosphate buffer (100 mM, pH 6) containing 0.1% (w/v) 3-trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP). A Bruker AV-400 MHz spectrometer operating at a resonance frequency of

399.427 MHz was used for the HR-MAS NMR experiments with a 4 mm HR-MAS dual inverse $^{1}H/^{13}C$ probe with magic angle gradient. A spinning frequency of 4 kHz and a temperature of 277 K was used during data acquisition.

A rotor synchronized Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence with water suppression was used to obtain one-dimensional ^{1}H HR-MAS spectra.[21] The one-dimensional spectra were collected applying 256 transients, a spectral width of 8000 Hz, data size of 16K point, an acquisition time of 2 seconds and a relaxation delay of 2 seconds. The free induction decays (FIDs) were exponentially weighted with a line broadening of 1 Hz. TOPSPIN 3.5 (Bruker Analytische Messtechnik, Germany) was used to phase the spectra manually and to perform automatically baseline correction. Gradient-enhanced two-dimensional ^{1}H-^{1}H COSY was obtained to confirm signal assignment.

Multivariate analysis

AMIX (version 3.8.7, BrukerBioSpin) is used to generate bucket tables for each time-point from the one-dimensional spectra excluding the region between 4.20 – 6.00 ppm to remove the larger water signal. The one-dimensional CPMG spectra were normalized to the total intensity and binned into buckets of 0.04 ppm. The data were mean-centred and scaled using the Pareto method in SIMCA software package (version 14.0, Umetrics, Umeå, Sweden). Supervised orthogonal partial least squares discriminant analysis (OPLS-DA) was performed on the data using the SIMCA software.

Quantification of the metabolites

The eighteen metabolites from our earlier metabolomics study of these mutants (Chapter 3) were quantified using the Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada). The known concentration of the reference peak of TSP was used to determine the concentration of the eighteen biomarkers. Metabolite concentrations are represented as means ± standard error. Student's t-test analysis of the NMR quantification results was performed with OriginPro 2016 (Northampton, USA).

4.4 Results and Discussion

Circadian rhythm of amino acids, proteins, sugars and starch

The functional pathways underlying the enhanced growth characteristics of the VP16-02-003 mutant and the VP16-05-014 mutant were investigated in this study throughout the circadian cycle. Prior to metabolic profiling, the circadian rhythms for the free amino acids, proteins, soluble sugars and starch are assayed (Figure 4.3).

During the dark period, amino acids are degraded.[22] In the wild-type Col-0, the free amino acid concentration shows an increasing trend during the light period and their levels dropped during the night period as expected (Figure 4.3A). Both mutants followed this pattern, and metabolite concentrations in VP16-02-003 were higher than for the Col-0. The concentration of free amino acids for the VP16-05-014 mutant was low at the beginning of the circadian cycle and very similar to the Col-0 in the dark period. The concentration of proteins remains constant throughout the light/dark cycle except for a dip at the beginning of the dark period (Figure 4.3B) and both mutants and wild-type follow the same circadian

rhythm. However, the level of proteins is lower throughout the cycle for the VP16-02-003 and VP16-05-014 mutant relative to Col-0.

In photosynthesis, light drives the assimilation of CO_2 into sugars. These sugars are used for several primary processes in *Arabidopsis thaliana*. Starch can be used in the dark period when light is not available.[23,24] For both sugars and starch, we expect a rhythm where the compound has risen to a maximum at the end of the light period and reduction takes place during the dark period.[23,24] The concentration of sugars in the *Arabidopsis* Col-0 has a peak at 12 hours and reduced during the dark period (Figure 4.3C). The VP16-02-003 mutant follows the same rhythm with a lower concentration throughout the whole cycle. The concentration of soluble sugars is constant for the VP16-05-014 throughout the light/dark cycle. In Col-0, the starch level elevated during the light period and dropped during the dark period (Figure 4.3D). The starch content of the VP16-02-003 mutant was constant during the whole period at a lower level in contrast to Col-0. VP16-05-014 follows the same rhythm as Col-0 but with a lower concentration. Most probably, soluble sugars and starch are not the most favourable carbon storage possibility in these mutants. Organic acids, like fumaric acid and malic acid, can be used as an alternative carbon storage in *Arabidopsis* plants.[25-27] These organic acids have been studied using metabolic profiling as described in the following section.

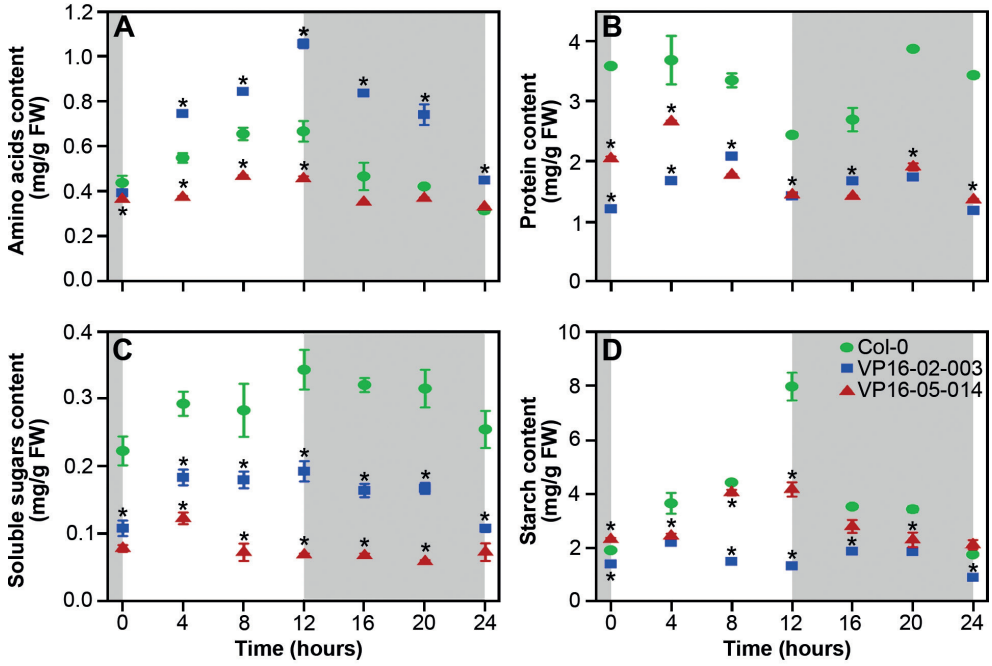


Figure 4.3: Changes in concentration of free amino acids (A), proteins (B), soluble sugars (C) and starch (D) throughout the circadian cycle for Arabidopsis thaliana Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲) in mg/g fresh weight. The results are given as mean ± SEM. Statistical analysis comparing groups was performed by Student's t-test ($p < 0.05$).*

Metabolic profiling of mutants at different time-points throughout the light/dark cycle using HR-MAS NMR

The metabolic profiles have been obtained from intact leaves of *Arabidopsis thaliana* at different time-points throughout the circadian cycle for the wild-type Col-0 and for both enhanced growth mutants. Figure 4.4 shows representative one-dimensional ^{1}H HR-MAS NMR spectra for Col-0, VP16-02-003 and VP16-05-014 at t = 12 hours.

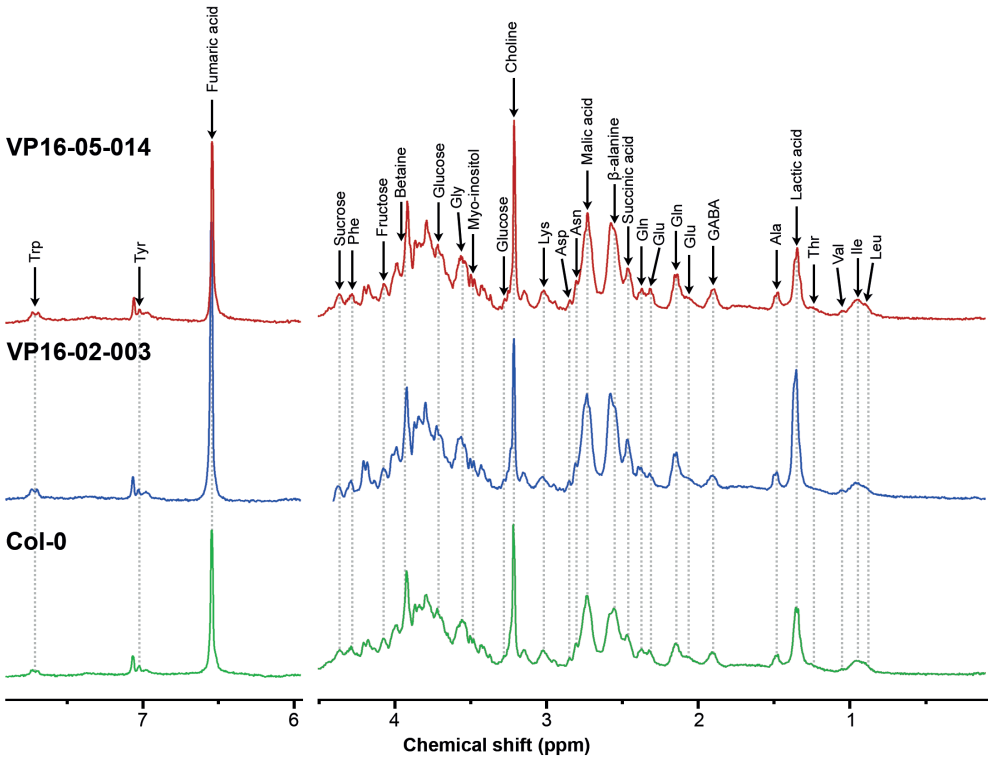


Figure 4.4: One-dimensional ^{1}H CPMG NMR spectra for Arabidopsis thaliana Col-0 (bottom panels), VP16-02-003 (middle panels) and VP16-05-014 (top panels) obtained from intact rosette leaf harvested at t = 12 hours into the light/dark cycle. Main metabolites have been assigned in the spectra.

The metabolic profiles for every time-point obtained by ^{1}H HR-MAS NMR were investigated by multivariate analysis of the bucket-reduced ^{1}H NMR data. Orthogonal partial least square discriminant analysis (OPLS-DA) was carried out to examine the variation in metabolic profile between two mutants and the wild-type Col-0 throughout the light/dark cycle (Figure 4.5). In the OPLS-DA model, this is realized by a separation of the orthogonal and predictive component of the replicates collected from the Col-0, VP16-02-003 and VP16-05-014. The predictive component is correlated with the differences between the Col-0 and mutant classes, while the orthogonal component is uncorrelated with these differences. Hence, in the OPLS-DA score plots, this leads to clustering and separation of the replicates belonging to each class at every time-point of the light/dark cycle.

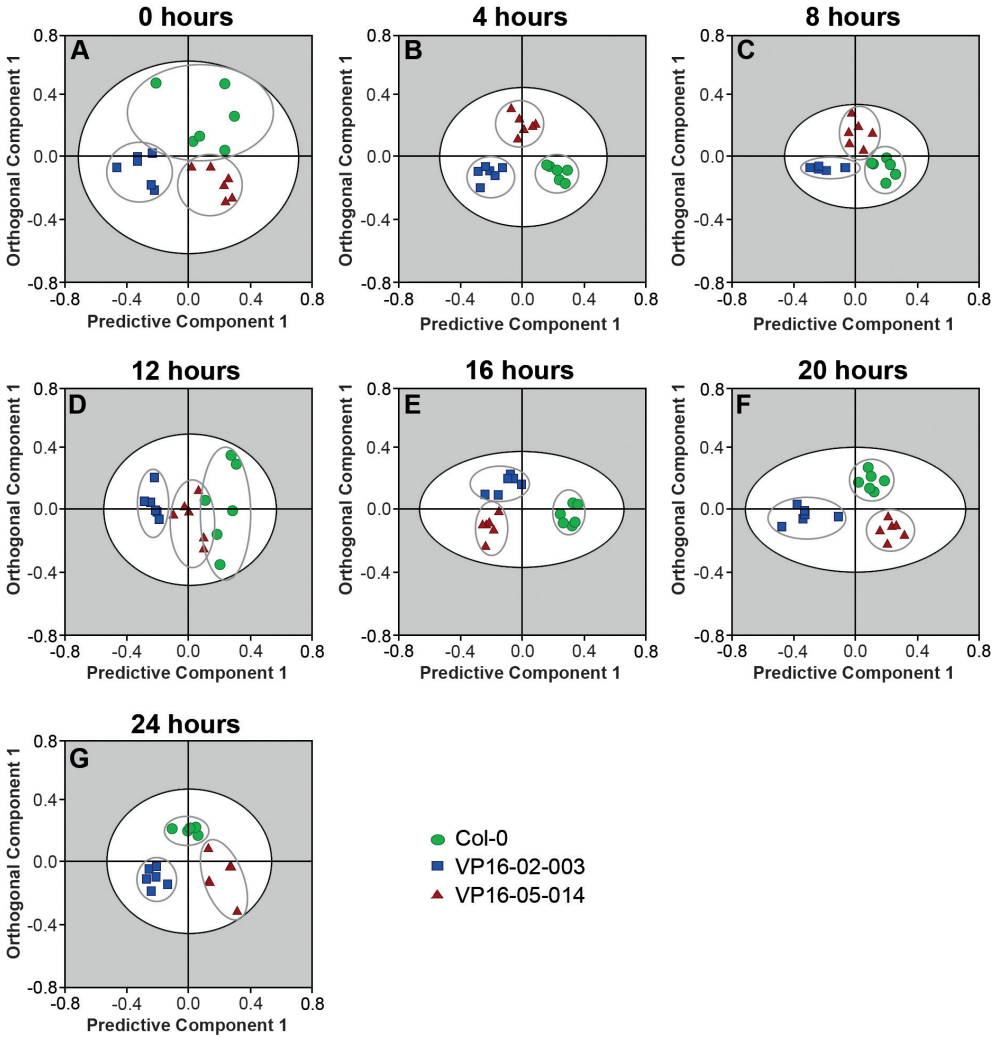


Figure 4.5: Orthogonal partial least square-discriminant analysis (OPLS-DA) score plots of the metabolite profiles derived from the intact leaves of Arabidopsis thaliana wild-type Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲) at the seven different time-points of harvesting throughout the light/dark cycle (see Figure 4.2).

In an earlier metabolomics study of the VP16-02-003 and VP16-05-014, eighteen biomarkers were identified at t = 6 hours after the start of the light period (Chapter 3). The circadian rhythm of these biomarkers is followed in this study. The primary metabolites that can be observed include the organic acids fumaric acid, malic acid and lactic acid, the sugars fructose and glucose, and the precursor of cell wall components choline (Figure 4.6) and the secondary metabolites myo-inositol, a sugar alcohol, and the organic osmolyte betaine. Also, 10 free amino acids can be resolved with NMR (Figure 4.7).

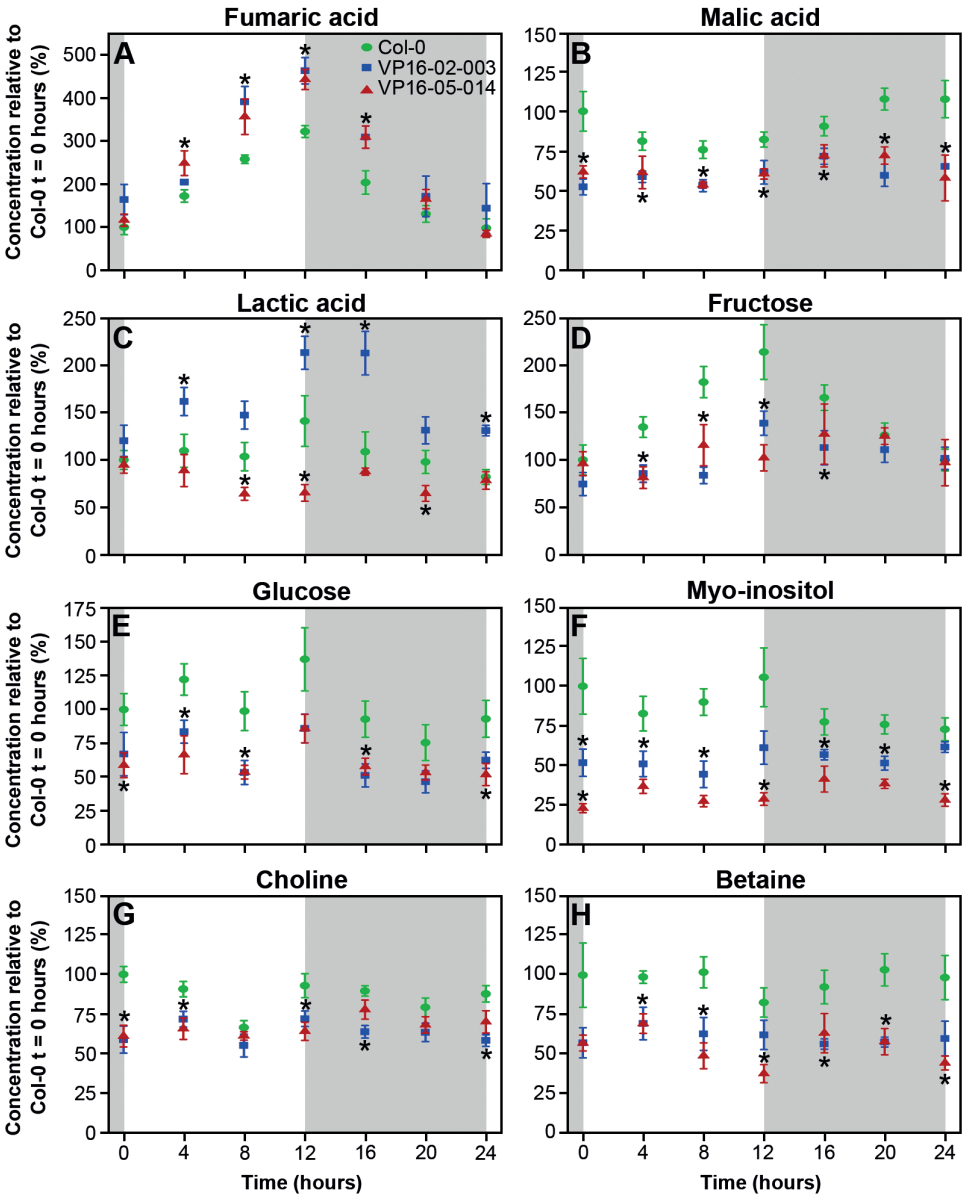


Figure 4.6: Metabolite concentrations during the circadian cycle of fumaric acid (A), malic acid (B), lactic acid (C), fructose (D), glucose (E), myo-inositol (F), choline (G) and betaine (H) in Arabidopsis thaliana Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲). Means ± SEM of 6 biological replicates is shown. Concentrations are expressed relative to the concentration of Col-0 at t = 0 hours. Statistical analysis comparing groups was performed by Student's t-test ($p < 0.05$). When two data points overlap, an asterisk indicates that both pass the Student's t-test.*

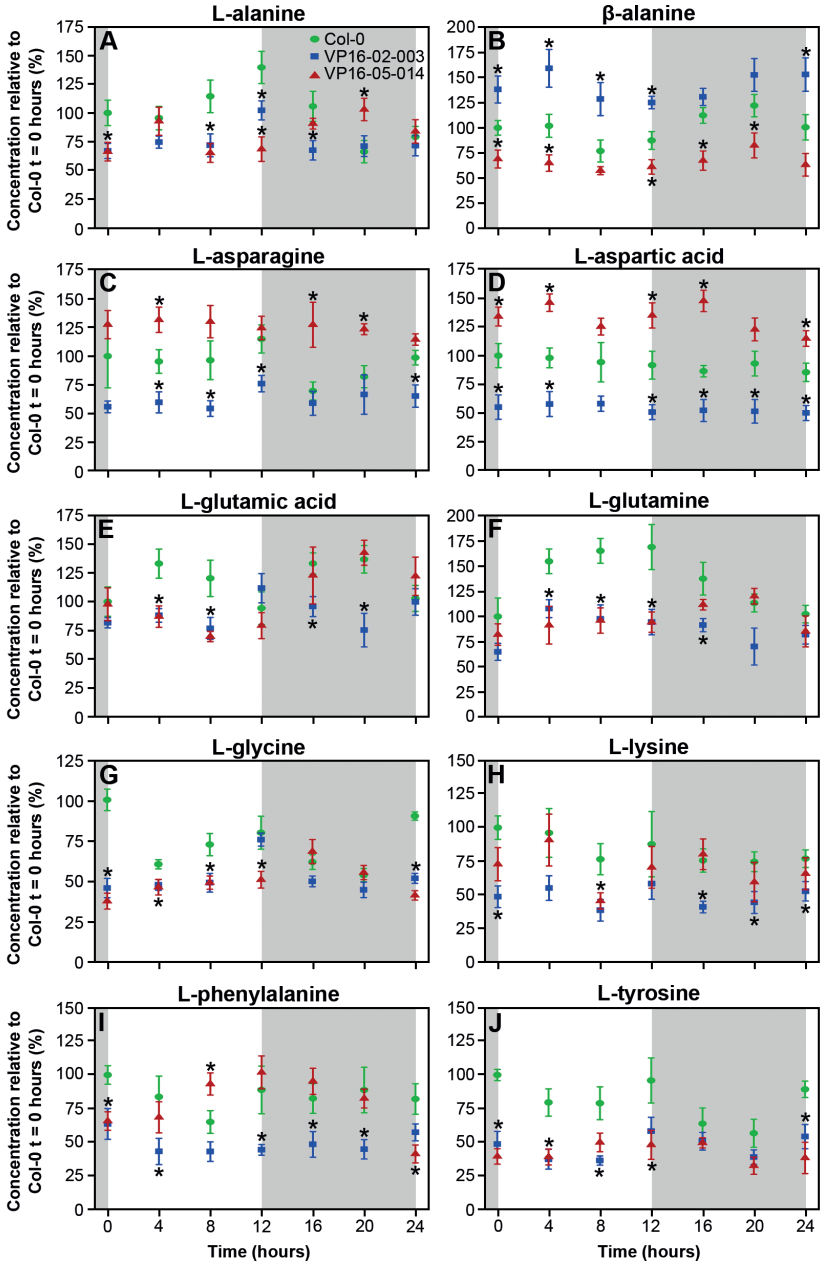


Figure 4.7: Concentration of the free amino acids L-alanine (A), β-alanine (B), L-asparagine (C), L-aspartic acid (D), L-glutamic acid (E), L-glutamine (F), L-glycine (G), L-lysine (H), L-phenylalanine (I) and L-tyrosine (J) throughout the circadian cycle in Arabidopsis thaliana Col-0 (●), VP16-02-003 (■) and VP16-05-014 (▲). Concentrations are expressed relative to the concentration of Col-0 at t = 0 hours. Statistical analysis comparing groups was performed by Student's t-test ($p < 0.05$).*

Characterization of metabolic rhythms throughout the circadian cycle

Fixation of CO_2 during photosynthesis is resulting in different photoassimilates, such as organic acids, sugars and starch. The circadian rhythms of the concentration of the organic acids fumaric acid, malic acid and lactic acid are shown in Figure 4.6A-C. In *Arabidopsis*, fumaric acid is generally considered the major form of fixed carbon and accumulates in the mitochondria and in the cytosol.[26-28] Figure 4.6A and 4.6B show that the concentrations of fumaric acid and malic acid increase during the light to a maximum at the end of the light period and decrease during the dark period in wild-type Col-0 and for both mutants.[7,27] The rhythm of the malic acid concentration in VP16-02-003 and VP16-05-014 remains the same in comparison to the wild-type Col-0, however with a lower concentration. In our earlier study, we found that the concentration of lactic acid is high at the end of the light period and dropped during the dark period in *Arabidopsis thaliana* Col-0.[7] VP16-02-003 follows this rhythm but at a higher concentration, while VP16-05-014 has a lower lactic acid concentration in comparison to Col-0 (Figure 4.6C).

Sugars can also be an intermediate of carbon fixation and play an important role in the circadian rhythms, mainly by regulation of diurnal genes. The photosynthetic carbon fixation during the light period partially produces sugars that can be used during the dark period.[22] Figure 4.6D and 4.6E show the concentration of fructose and glucose during the light/dark cycle. The concentration of fructose and glucose increases during the light period and declines during the dark period in *Arabidopsis* Col-0 and both mutants. VP16-02-003 and VP16-05-014 have lower concentrations throughout the whole light/dark cycle than for the wild-type. Apparently both mutants VP16-02-003 and VP16-05-014 prefer carbon storage into organic acids instead of sugars.

Myo-inositol is synthesized from glucose and has diverse biological roles including phosphate storage, cell wall biogenesis and stress tolerance.[29] Figure 4.6F shows the circadian rhythm of myo-inositol for the wild-type Col-0 and the two mutants VP16-02-003 and VP16-05-014. The variation of the concentration of myo-inositol is very minor over the entire light/dark cycle in Col-0 and the VP16-02-003 and the VP16-05-014 mutant, which is in line with mass spectrometry data collected by Gibon *et al.* for the wild-type.[30] Both mutants have a lower concentration of myo-inositol throughout the whole cycle compared to Col-0.

Choline is an essential metabolite which is needed to synthesize membrane phospholipids in plants. Choline can be oxidized to the organic osmolyte betaine.[31,32] For Col-0, the levels of choline decrease during the light period and increase during the dark period (Figure 4.6G). The VP16-02-003 and VP16-05-014 mutants follow the same rhythm at a lower choline level than for Col-0. The concentration of betaine decreases during the light period and increases during the dark period (Figure 4.6H) in wild-type *Arabidopsis* Col-0, the VP16-02-003 and the VP16-05-014 mutant. The level of betaine is reduced for both mutants compared to Col-0.

Figure 4.7 shows the circadian rhythm of the concentrations of ten free amino acids for the wild-type Col-0 and the VP16-02-003 and VP16-05-014 mutants. The metabolic rhythm of the free amino acids does not differ between the mutants and the wild-type Col-0 (Figure 4.7). The free amino acid alanine, glutamic acid, glutamine, glycine, lysine, phenylalanine

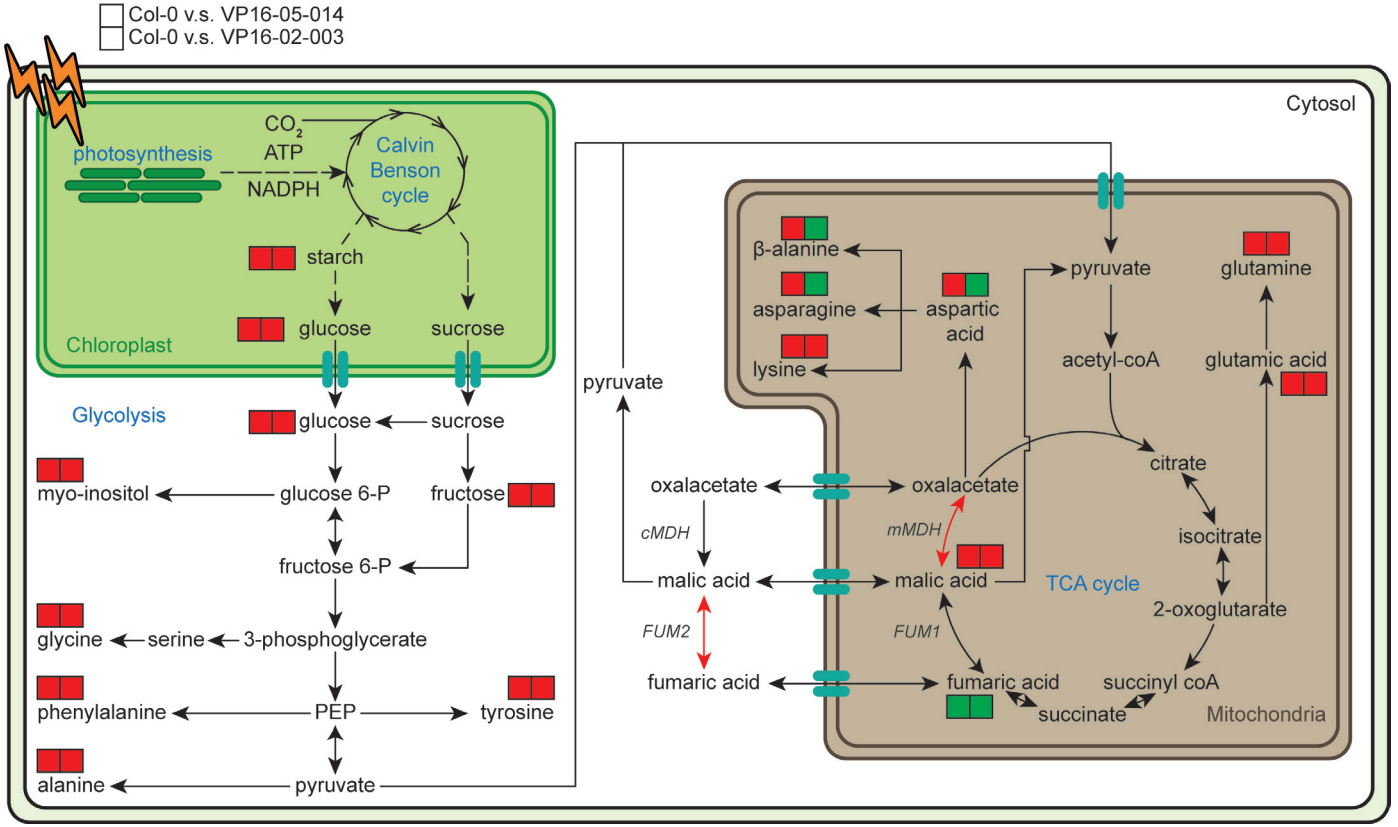


Figure 4.8: Pathway view of the central carbon metabolism of primary metabolites for the VP16-02-003 and VP16-05-014 mutant in comparison to Col-0. Colours indicates higher levels (green) or lower levels (red) of the primary metabolite. Dashed lines indicate multiple conversion steps.

and tyrosine have lower concentrations in both mutants in comparison to Col-0. The concentrations of aspartic acid, asparagine and β-alanine are higher in VP16-02-003 and lower in the VP16-05-014 mutant in comparison to Col-0.

Figure 4.8 summarizes the results of levels of the primary metabolites in comparison to Col-0 in the central carbon metabolism of *Arabidopsis*. What is remarkable is that only the concentration of fumaric acid increases for both mutants relative to Col-0, while the levels of the other primary metabolites are decreased for both mutants. The level of the amino acids aspartic acid, asparagine and β-alanine are upregulated in the VP16-05-014 while these amino acids are downregulated in the VP16-02-003 mutant in comparison to Col-0. Other detected free amino acids have lower concentration for both mutants in contrast to the wild-type. The lower concentrations of sugars, myo-inositol and free amino acids for the mutants compared to Col-0 were attributed to an impaired stress response in the previous chapter 3.

To better understand the high accumulation of fumaric acid, the TCA cycle has been studied in more detail by combining the metabolic data with the mRNA expression levels from the RNA-seq data.[6] The expression level of mitochondrial malate dehydrogenase 1 (mMDH1, AT1G53240) is reduced for both the VP16-02-003 and the VP16-05-014 mutant compared to Col-0. Malic acid can be converted to pyruvate in the mitochondria and in the cytosol. This conversion is facilitated by high concentrations of fumaric acid and may lead to increased fluxes through acetyl-coA back into the TCA cycle.[25]

Fumaric acid and malic acid can be transported to the cytosol and can be interconverted in each other by cytosolic fumarase 2 (FUM2, AT5G50950).[33] The mRNA expression level of FUM2 is reduced for both mutants in comparison to Col-0. The expression level of FUM2 for the VP16-02-003 mutant (logFC = -1.03) is even lower than for the VP16-05-014 mutant (logFC = -0.54). The natural *Arabidopsis* accession Rsch-4 has also reduced expression of FUM2 due to an insertion/deletion polymorphism in the promotor of the FUM2 gene and accumulates high levels of fumaric acid.[33,34] Next to the similarity in high levels of fumaric acid and reduced expression of FUM2, the rosette area of the *Arabidopsis* Rsch-4 is also larger in comparison to Col-0.[35]

The relation between the high levels of fumaric acid and the occurrence of growth vigour is not yet clear. Earlier occurrences of growth vigour in *Arabidopsis* have been related to ploidy and hybridity effects subject to circadian regulation.[36] In addition, epigenetic regulation can lead to growth enhancement.[37] An example is the *Arabidopsis msh1* mutant where the nuclear-encoded MutS HOMOLOGUE 1 (MSH1) gene is downregulated which triggers nuclear epigenetic reprogramming leading to enhanced growth vigour.[37] The zinc finger artificial transcription factors method is a very recent epigenetic reprogramming method to enhance growth.[5,6]

For better understanding, the relation between fumaric acid accumulation and growth vigour, and how primary metabolite concentrations are balanced in the interplay between photosynthetic source and utilization sinks leading to growth, examination of the enzyme activities of malate dehydrogenase and fumarase may lead to underpinning

the underlying mechanisms of conversion of fumaric acid and malic acid.[38] Specific phenotypes can be examined by fluxomics where the metabolic fluxes are studied as the interplay of gene expression, protein concentration, kinetics, regulation and metabolite concentrations.[39] In the general outlook and future outlook (Chapter 5), the fluxomics approach will be described in more detail.

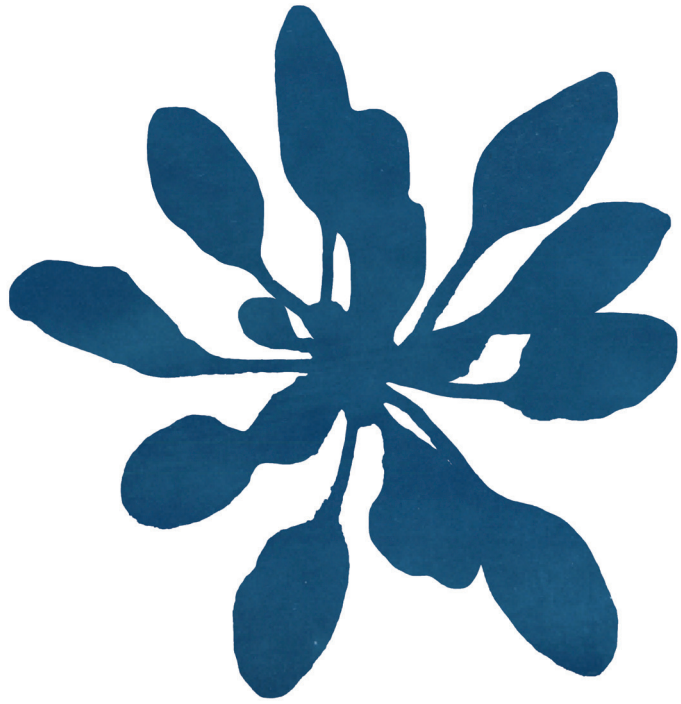
4.5 Conclusion

In this study, we investigated the circadian rhythm of the eighteen earlier identified biomarker for the enhanced growth characteristics phenotype of the VP16-02-003 and VP16-05-014 mutant. HR-MAS NMR was used to obtain the metabolic profile throughout the light/dark cycle for *Arabidopsis thaliana* wild-type Col-0 and the VP16-02-003 and VP16-05-014 mutant. The metabolic rhythm of the eighteen biomarkers is not changed, while the concentrations differ throughout the whole light/dark cycle. Since the clock functional periodicity is independent of the cellular complexity and growth-defence trade-off, the results contribute to converging evidence that it may not be necessary to - upstream - alter the circadian clock when the goal is to achieve enhanced growth characteristics, and that - downstream - phenotypic engineering of sinks and bottlenecks leading to growth may be more effective in a multifactorial context that can be altered by whole genome reprogramming.

4.6 References

[1] C. Mba et al., Agriculture & Food Security 2012, 1, 7–17.
[2] A. Kumar et al., OMICS 2015, 19, 581–601.
[3] V. S. Kamburova et al., International Journal of Agronomy 2017, 2017, 15.
[4] T. Gaj et al., Trends in Biotechnology 2013, 31, 397–405.
[5] N. van Tol, B. J. van der Zaal, Plant Science 2014, 225, 58–67.
[6] N. van Tol et al., PLoS One 2017, 12, e0174236.
[7] D. Augustijn et al., PLoS One 2016, 11, e0163258.
[8] S. Caretto et al., Int. J. Mol. Sci. 2015, 16, 26378–26394.
[9] B. Huot et al., Molecular Plant 2014, 7, 1267–1287.
[10] T. Zust, A. A. Agrawal, Annu Rev Plant Biol 2017, 68, 513–534.
[11] S. Barak et al., Trends in Plant Science 2000, 5, 517–522.
[12] K. Greenham, C. R. McClung, Nat Rev Genet n.d., 16, 598–610.
[13] J. D. Thomas et al., Annu Rev Biophys Biomol Struct 2004, 33, 75–93.
[14] H. Huang, D. A. Nusinow, Trends in Genetics 2016, 32, 674–686.
[15] N. Nakamichi, Plant Cell Physiol 2011, 52, 1709–1718.
[16] J. S. Shim et al., Plant Physiology 2017, 173, 5–15.
[17] S. Gepstein, K. V. Thimann, Plant Physiology 1981, 68, 349–354.
[18] Y. Gibon et al., Plant Cell 2004, 16, 3304–3325.
[19] M. DuBois et al., Anal. Chem. 1956, 28, 350–356.
[20] A. M. Smith, S. C. Zeeman, Nat. Protocols 2006, 1, 1342–1345.
[21] S. Meiboom, D. Gill, Review of Scientific Instruments 1958, 29, 688–691.
[22] O. E. Bläsing et al., Plant Cell 2005, 17, 3257–3281.
[23] M. J. Haydon et al., Nature 2013, 502, 689–692.
[24] M. Seki et al., Sci. Rep. 2017, 7, 8305.
[25] A. U. Igamberdiev, A. T. Eprintsev, Front. Plant Sci. 2016, 7, 513.
[26] M. B. Zell et al., Plant Physiology 2010, 152, 1251–1262.

[27] D. W. Chia et al., Planta 2000, 211, 743–751.
[28] H. Fahnenstich et al., Plant Physiology 2007, 145, 640–652.
[29] P. H. Meng et al., PLoS One 2009, 4, e7364.
[30] Y. Gibon et al., Genome Biol. 2006, 7, R76.
[31] A. Sakamoto, N. Murata, Plant Cell Environ 2002, 25, 163–171.
[32] J. Huang et al., Plant Physiol Biochem 2008, 46, 647–654.
[33] J. Weiszmann et al., 2017, DOI 10.1101/168617.
[34] D. Riewe et al., The Plant Journal 2016, 88, 826–838.
[35] A. Camargo et al., PLoS One 2014, 9, e96889.
[36] M. Miller et al., G3: Genes|Genomes|Genetics 2012, 2, 505–513.
[37] K. S. Virdi et al., Nature Communications n.d., 6, 6386 EP –.
[38] M. Stitt, Y. Gibon, Trends in Plant Science 2014, 19, 256–265.
[39] G. Winter, J. O. Krömer, Environ Microbiol 2013, 15, 1901–1916.



General discussion and future outlook

To meet the increasing demand for biomass resources for both bioenergy as food, the biomass yield of agriculture products needs to be improved. Agriculture products with increased biomass yield and with a minimal environmental footprint are called smart crops.[1,2] Smart crops can be developed by genome editing methods.[3-6] In the PhD thesis, I examined the phenotype of two phenotypically engineered *Arabidopsis thaliana* mutants with increased growth characteristics using a systems biology approach combining omics technologies and bio-imaging methods to understand which pathways are underlying this larger rosette surface phenotype.

5.1 Understanding the phenotype of *Arabidopsis* mutants

A high-resolution magic angle spinning (HR-MAS) NMR-based approach is used to obtain the metabolic profile directly from the leaves of the *Arabidopsis thaliana* plant. The major advantages of HR-MAS NMR over liquid-state NMR is that there is no extraction step necessary which can lead to the loss of signals from non-soluble metabolites. It is also a non-destructive method, which makes it possible to use the sample for other experiments such as transcriptomics analysis.[7,8] To our knowledge, this is the first time that HR-MAS NMR was applied to resolve metabolic profiles in *Arabidopsis thaliana* leaves, in the native state. Combined with multivariate analysis, HR-MAS NMR-based metabolomics is a powerful tool to investigate the phenotype of plants. It can link the gene regulatory network and the functional physical structure, which is considered as highly complex, using an axiomatic design approach (Figure 5.1A).[9]

Evolution and development is an interplay between variation at the basis and selection from the higher levels. Metabolic engineering cannot proceed bottom-up only, it needs engineering of a selection criterion to steer the system in the desired direction. Reducing the complexity for highly complicated biological systems gives insight in the system range,

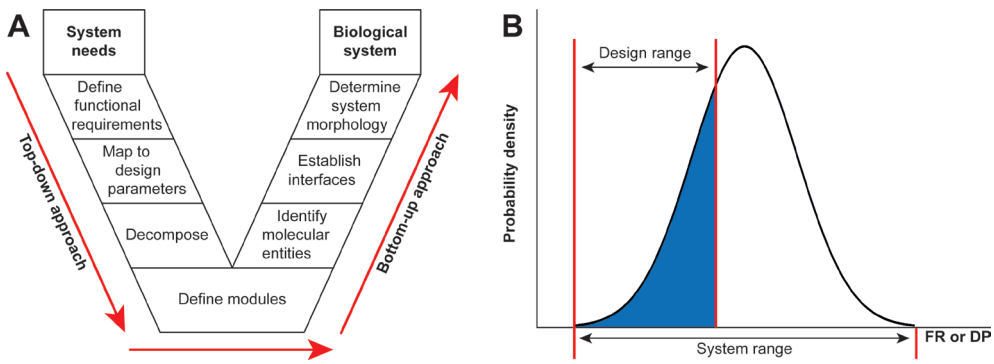


Figure 5.1: V-model can be used to understand the molecular interactions leading to the phenotype of plants. The first step is to define the systems need and to define the functional requirements. The functional requirements are mapped to the design parameters such as molecules, cells etc to decompose phenotype. When the physical modules are clear, they can be put together to have a biologically functioning system by identifying the molecular entities, establish the interfaces and finally to determine the system morphology.[4] B) The system range represents the different phenotype, which is a probability density function. The design range is the set of phenotypes which can be obtained.[4]

the range of phenotypes that can be realized by variation at the molecular level, and how functional requirements, such as a proper balance between sources and sinks, determine the specific range of phenotypes that fulfil the requirements of the biological design, the design range imposed by selection from the higher levels in the biological hierarchy (Figure 5.1B).

5.2 *Arabidopsis* mutants with enhanced growth characteristics

In this thesis, I apply the HR-MAS NMR toolbox combined with systems biology to investigate the *Arabidopsis thaliana* mutants with enhanced growth characteristics to understand which pathways are leading to enhanced growth. The results show that the VP16-02-003 and VP16-05-014 mutants have a reduced ability to defend against stresses. The growth-defence trade-off indicates that the energy which is not used for defence is available for enhanced growth.

The disadvantage of this reduced ability to defend against environmental stress is that the plants need to be cultivated in a protected environment. A protected environment can be achieved with controlled-environmental agriculture (CEA) technologies, like greenhouses.[10] Greenhouses take up a lot of the available space for agriculture. A way to solve this is to make use of vertical farming (Figure 5.2). In vertical farming, plants are cultivated in stacked layers and in an environment where light, temperature and humidity can be regulated.[10,11]

Figure 5.2: An example of vertical farming with different levels to grow plants (© Adobe Stock).

5.3 *Arabidopsis* Low Chlorophyll Fluorescence 1 mutant

In addition to the studies presented in this thesis, I applied the HR-MAS NMR-based metabolomics workflow to the *Arabidopsis* Low Chlorophyll Fluorescence 1 (LCF1) mutant.[12] Like the VP16-02-003 and VP16-05-014 mutants, the LCF1 mutant is obtained with artificial zinc-finger genome interrogation. LCF1 has a high operating efficiency of photosystem II (ϕPSII) and low chlorophyll fluorescence. From an early study, we know that the PSII/PSI ratio is decreased in this mutant relative to Col-0. Also, the rosette surface area and biomass yield are reduced, but the starch content per gram fresh weight is increased compared to the wild-type.[12] The LCF1 mutant expresses a truncated MORC2 protein which encodes an ATPase of the Microrchidia (MORC) family. This gene is involved in large-scale epigenetic transcriptional regulation. There is converging evidence that the truncated protein forms the basis of the LCF1 phenotype. To better understand the alterations in the pathways leading to the phenotype of this plant, the HR-MAS NMR-based metabolic workflow was applied on the leaves of the LCF1 mutant at seven different time-points throughout the light/dark cycle. The results of this study will be published separately.

Figure 5.3 shows the score plots of the multivariate analysis of the *Arabidopsis* wild-type Col-0 and the LCF1 mutant at t = 4 hours into the light period. The three component PCA model explains 84.4% of the separation (Q^2 = 0.6). For the OPLS-DA model, the R^2X, R^2Y and Q^2 (cum) were 0.822, 0.952 and 0.733, respectively. Both models have a good quality and an accurate prediction. The sample classes are clustered together in both the PCA and the OPLS-DA score plot.

Both PCA and OPLS-DA models were constructed for every time-point in the light/dark cycle. The OPLS-DA model is used to identify sixteen biomarker candidates and quantified the metabolite levels of these metabolites. This includes the organic acids fumaric acid, malic acid and lactic acid, the sugar fructose, the cofactor NADPH, the organic osmolyte

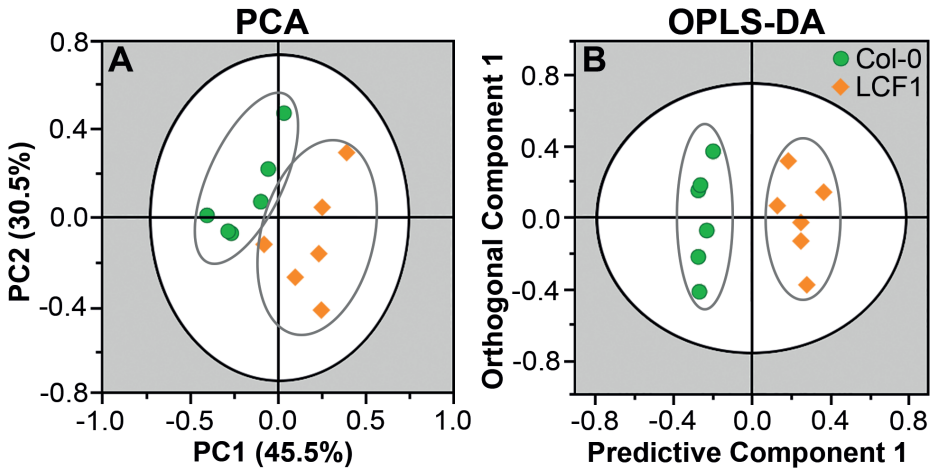


Figure 5.3: The score plots for the PCA (A) and OPLS-DA (B) analysis from the metabolic profiles obtained from the whole Arabidopsis thaliana leaves of Col-0 (●) and LCF1 (◆) at t = 4 hours into the light period.

betaine, the amino acids L-aspartic acid, L-glutamic acid, L-leucine, L-lysine, L-phenylalanine, L-threonine, L-tyrosine, L-valine and the non-essential amino acid GABA and the non-proteinogenic amino acid ornithine. The next step is to perform literature research and pathway modelling to investigate the function of these biomarkers and relate them to the phenotype of the LCF1 mutant.

5.4 Future outlook

In the future, the HR-MAS NMR-based metabolomics workflow can be used to investigate other *Arabidopsis thaliana* mutants. In addition, it can be applied to smart crops to probe the primary metabolites and selected secondary metabolites. It is also possible to make some improvements in the HR-MAS NMR technology for example by improving the resolution, localized HR-MAS NMR and HR-MAS NMR on living plants. In addition, the results of metabolomics can be supplemented with information about altered metabolic steady-states in the system using fluxomics. Also, the results from the metabolomics can be integrated with other omics technologies in a multi-omics approach.

Crops for bioenergy

In my PhD project, I investigated *Arabidopsis thaliana* mutants obtained using zinc-finger artificial transcription factor (ZF-ATF)-mediated whole genome interrogation and used HR-MAS NMR-based metabolomics to understand the underlying pathways resulting in the phenotypes of the plants. For instance, several plants are interesting to use as biological resources for bioenergy. Traditional cereal crops like maize (*Zea mays*) and sorghum (*Sorghum bicolor*) are interesting for starch-based ethanol production. Also, sugar-producing plants like sugarcane (*Saccharum officinarum*) and sugar beet (*Beta vulgaris*) can be used for direct fermentation of ethanol. But there are also crops dedicated to bioenergy production, such as Switchgrass (*Panicum virgatum*) and woody plants, including hybrid poplar, willow and pines. For biodiesel, soybean (*Glycine max*), canola (*Brassica napus*) and sunflower (*Helianthus annuus*) are interesting crops. Alternative bioenergy plants like grasses, trees and green algae are interesting to investigate using the systems biology approach.[13]

Improvement of the HR-MAS NMR technology

Among improvements to be made in the HR-MAS NMR technology, an obvious one is to use an NMR spectrometer with a higher magnetic field strength to improve the sensitivity and resolution of the NMR spectra. This can result in the observation of lower abundant metabolites at higher field strength. Soon we will have capabilities to perform HR-MAS NMR at 1.2 GHz, which will improve the range and resolution of the method relative to the current practice in this thesis, where experiments were performed at 400 MHz.[14]

It also can be interesting to study specific structures of the leaves, such as the veins, lamina or the petiole or other parts of the plant. Recently, Sarou-Kanian *et al.* developed a new method using ^{1}H HR-MAS slice localized spectroscopy (SLS) and HR-MAS chemical shift imaging (CSI) to determine the distribution of metabolites along the anteroposterior axis of *Drosophila melanogaster*.[15] Here a MAS probe coupled with a three axes gradient system is used, together with pulse sequences for SLS and CSI. The same technology can be applied for the investigation of plant material.

Fluxomics and multi-omics integration

Metabolomics provides a snapshot of the metabolic status of a sample at a specific time. For most enzymes involved in metabolism, knowledge about the *in vivo* kinetics is necessary to predict metabolic fluxes. Metabolic fluxes are the result of the interplay of gene expression, protein concentration, protein kinetics and regulation, and depend on metabolite concentrations (red arrows in Figure 5.4). Metabolic flux analysis, also called fluxomics, can be used to determine the metabolic reaction rates. Fluxomics can thus help to understand complex metabolic pathways and their regulation for characterisation of the phenotype of the plant.[16,17] Fluxomics can be done by introducing a ^{13}C-labelled precursor into the metabolic network or by supplying $^{13}CO_2$ and follow the redistribution of the label into other metabolites by either NMR or mass spectrometry.[18,19] The redistribution can be followed throughout time during dynamic labelling or after reaching steady-state in a steady-state labelling approach.[19] In the current fluxomics protocols, an extraction step has to be performed which has the disadvantage of losing components during preparation. It can thus be interesting to develop an HR-MAS NMR-based fluxomics approach which is not available at the moment.

In a multi-omics approach, the results from the various omics technologies, such as genomics, transcriptomics, proteomics, metabolomics and fluxomics, are integrated to unravel the complexity of a biological system (Figure 5.4).[20-22] A major practical challenge of multi-omics is to handle different data formats and the high data dimensionality property of each data set. To integrate the different information layers, bioinformatics tools are necessary to track the different components for every layer, such as genes, proteins and metabolites at the same time.[20]

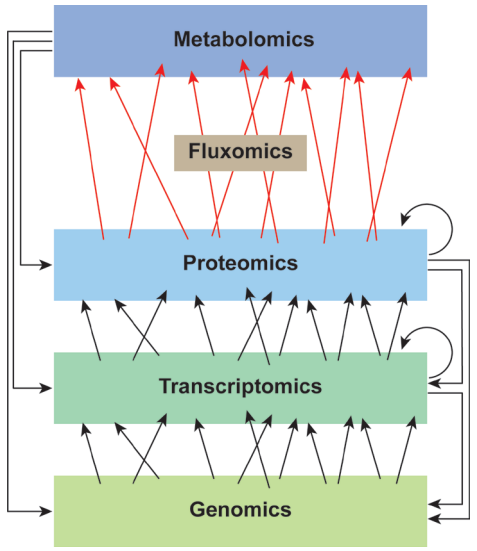


Figure 5.4: The results of genomics, transcriptomics, proteomics and metabolomics are integrated in a multi-omics approach to unravel the complexity of a biological system. In fluxomics (red arrows), the interplay between gene expression, protein concentration, protein kinetics and regulation and the metabolite concentrations are studied.

5.5 References

[1] C. Mba et al., Agriculture & Food Security 2012, 1, 7–17.
[2] A. Kumar et al., OMICS 2015, 19, 581–601.
[3] N. van Tol, B. J. van der Zaal, Plant Science 2014, 225, 58–67.
[4] T. K. Mohanta et al., Genes (Basel) 2017, 8, 399.
[5] B. I. Laufer, S. M. Singh, Epigenetics & Chromatin 2015 8:1 2015, 8, 34.
[6] T. Gaj et al., Trends in Biotechnology 2013, 31, 397–405.
[7] A. Wong, C. Lucas-Torres, in New Developments in NMR, Cambridge, 2018, pp. 133–150.
[8] K. Shet et al., NMR Biomed 2012, 25, 538–544.
[9] J. D. Thomas et al., Annu Rev Biophys Biomol Struct 2004, 33, 75–93.
[10] K. Benke, B. Tomkins, Sustainability: Science, Practice and Policy 2017, 13, 13–26.
[11] D. Despommier, The Vertical Farm: Controlled Environment Agriculture Carried Out in Tall Buildings Would Create Greater Food Safety and Security for Large Urban Populations, 2011.
[12] N. van Tol et al., Sci. Rep. 2017, 7, 3314.
[13] J. S. Yuan et al., Trends in Plant Science 2008, 13, 421–429.
[14] H. C. Bertram et al., Anal. Chem. 2007, 79, 7110–7115.
[15] V. Sarou-Kanian et al., Sci. Rep. 2015, 5, 9872.
[16] C. Salon et al., Journal of Experimental Botany 2017, 68, 2083–2098.
[17] G. Winter, J. O. Krömer, Environ Microbiol 2013, 15, 1901–1916.
[18] K. Kölling et al., Plant Methods 2013, 9, 45–45.
[19] R. G. Ratcliffe, Y. Shachar-Hill, The Plant Journal 2006, 45, 490–511.
[20] M. Fondi, P. Lio, Microbiol Res 2015, 171, 52–64.
[21] M. Kim, I. Tagkopoulos, Mol. Omics 2018, 14, 8–25.
[22] M. D. Ritchie et al., Nat Rev Genet 2015, 16, 85–97.

Summary

Climate change is a challenge for both current and future generations. New biological resources have to be developed in order to meet the demand for energy as well as the demand for food for the rising world population. One way of doing this is to make use of so-called smart crops, as they have improved yields and a small environmental footprint. The development of smart crops can be accelerated by using genome editing methods like zinc finger artificial transcription factor-mediated whole genome interrogation to obtain novel mutant phenotypes. Model plant organisms, such as *Arabidopsis thaliana*, can contribute to the development of smart crops as they have been studied extensively by the research community. In this PhD thesis, two phenotypically engineered Arabidopsis mutants with enhanced growth characteristics have been studied.

The purpose of this PhD thesis is to unravel the pathways leading to the phenotype of *Arabidopsis thaliana* mutants from a systems biology perspective, by using metabolomics. In metabolomics, metabolites are studied both qualitatively and quantitively under specific growth conditions. Since metabolites are the end products of cellular processes, the metabolome is most closely related to the phenotype of plants. To obtain the metabolic profile directly from the leaves of *Arabidopsis*, high-resolution magic angle spinning (HR-MAS) NMR is used. Combined with multivariate analysis, biomarkers, related to the mutant phenotype, are determined. The biological interpretation of the biomarkers will be performed in a systems biology approach by using the available information from the literature to obtain a model to explain the mutant phenotype.

Chapter 1 contains a detailed explanation of metabolomics and the role of metabolomics in systems biology. The most common techniques to obtain metabolic profiles are mass spectrometry and liquid-state NMR. Extraction of metabolites from the sample is necessary for these techniques, which leads to the loss or degradation of metabolites during extraction. The extraction procedure can be eliminated by using HR-MAS NMR which enables the use of intact leaf material. It is, therefore, a powerful tool in plant metabolomics and it has been used for an increasingly wide range of applications. In this chapter, a HR-MAS NMR-based workflow is proposed and described, including used pulse sequences, data pre-processing and multivariate analysis.

Chapter 2 describes the metabolic profile that has been collected directly from the Arabidopsis thaliana leaves by using HR-MAS NMR. Wild-type *Arabidopsis thaliana* accession Columbia-0 (Col-0) has been studied throughout the light/dark cycle to examine primary metabolites and to determine their functional periodicity during the circadian cycle. Both one- and two-dimensional ^{1}H HR-MAS NMR spectra are used to identify and quantify primary metabolites including organic acids, sugars and amino acids. Multivariate analysis revealed modules of primary metabolites for which the concentrations vary during the circadian cycle.

In **Chapter 3**, the HR-MAS NMR-based workflow is used to obtain the metabolic profile

for the *Arabidopsis thaliana* mutant VP16-02-003 and VP16-05-014 with enhanced growth characteristics at a time-point in the middle of the light period. Eighteen identified metabolites showed significant alteration of the metabolite concentration between the mutants and the wild-type Col-0. A biological interpretation of the concentration levels of the metabolites in terms of a reduced ability of the plants to respond against stress is presented. The diminished defence response leads to an altered growth-defence trade-off for both mutants. The results in this chapter contribute to converging evidence that improved biomass yield can be obtained by changing the balance between growth and defence.

Chapter 4 describes how the eighteen biomarkers for the enhanced growth characteristics of both mutants can be studied throughout the light/dark cycle, as the photosynthetic machinery is under control of the circadian clock. The circadian rhythm is robust, since the variation of the metabolite concentrations is very similar for the wild-type Col-0, the VP16-02-003 and the VP16-05-014. In contrast, the overall metabolite concentrations, throughout the whole circadian cycle, are different for the three species. Hence the clock functional periodicity appears independent of the growth-defence trade-off. This indicates that downstream phenotypic engineering of sinks and bottlenecks leading to growth is more effective than altering the circadian clock upstream.

Finally, **Chapter 5** provides a general discussion of this work and a future outlook. In conclusion, HR-MAS NMR metabolomics combined with multivariate analysis is a powerful tool to understand the phenotype of plants and it can be used for *Arabidopsis* mutants or other crops which are interesting to use as biological resources for bioenergy. In fluxomics, the metabolic fluxes, resulting from the interplay of gene expression, protein concentration, protein kinetics and regulation and metabolite levels, can be used to interpreted complex metabolic pathways and the regulation for a better understanding of the phenotype of plants. Thus, when metabolomics is combined with other omics technologies, such as transcriptomics and fluxomics, the complexity of a biological system can be further unravelled.

Samenvatting

Klimaatverandering is een uitdaging voor zowel huidige als toekomstige generaties. Om ervoor te zorgen dat de groeiende wereldbevolking voldoende energie en voedsel tot haar beschikking heeft, moeten er nieuwe biologische hulpbronnen worden ontwikkeld. Dit kan bereikt worden door gebruik te maken van zogenaamde slimme gewassen, vanwege hun hogere biomassa opbrengst en hun minimale ecologische voetafdruk. De ontwikkeling van slimme gewassen kan worden versneld door gebruik te maken van genome editing methoden, zoals zinkvinger artificiële transcriptie factoren-gemedieerde genoom aanpassingen. Model plant organismen, zoals *Arabidopsis thaliana*, kunnen helpen bij de ontwikkeling van slimme gewassen, omdat er al uitgebreid onderzoek is gedaan naar deze plant dat veel achtergrondinformatie heeft opgeleverd. In dit proefschrift zijn twee *Arabidopsis* mutanten met een hogere biomassa accumulatie nader bestudeerd.

Het doel van dit proefschrift is om inzicht te verwerven in metabolische routes die leiden tot het fenotype van de *Arabidopsis thaliana* mutanten vanuit een systeem biologisch perspectief met behulp van metabolomics. In de metabolomics worden planten onder zeer specifieke omstandigheden bestudeerd, op zowel kwalitatieve als op kwantitatieve wijze. Aangezien metabolieten de eindproducten zijn van cellulaire processen, is het metaboloom het meest verwant aan het fenotype van planten. Om het metabole profiel direct af te leiden van de bladeren van *Arabidopsis*, is er gebruik gemaakt van hoge-resolutie “magic angle spinnig” (HR-MAS) NMR. In combinatie met multivariate analyse worden biomarkers vastgesteld die gerelateerd zijn aan het fenotype van de mutant. Voor de biologische interpretatie van de biomarkers wordt gebruik gemaakt van de systeem biologie-methode, waarbij beschikbare informatie uit de literatuur gebruikt wordt om het fenotype van de mutant te kunnen verklaren door middel van een model.

Hoofdstuk 1 bevat een gedetailleerde uitleg van de metabolomics en de rol van metabolomics in de systeem biologie. De meest gebruikte technieken om het metabole profiel te bepalen zijn massaspectrometrie en vloeistof NMR. Voor beide technieken is extractie van de metabolieten uit de bladeren nodig, wat kan leiden tot verlies of afbraak van de metabolieten tijdens de extractie. De extractieprocedure kan worden geëlimineerd door gebruik te maken van HR-MAS NMR, waarbij het bestuderen van intact materiaal mogelijk is. Dit hoofdstuk bevat een gedetailleerd voorstel voor een HR-MAS NMR-workflow, inclusief de gebruikte pulssequenties, data-preprocessing en multivariate analyse. HR-MAS NMR is een krachtig hulpmiddel in de planten metabolomics, wat gebruikt kan worden voor een breed scala van toepassingen. Deze toepassingen worden tevens samengevat in dit hoofdstuk.

In **Hoofdstuk 2** wordt het metabole profiel beschreven van *Arabidopsis thaliana* bladeren, verkregen met behulp van HR-MAS NMR. Het wild-type, *Arabidopsis thaliana* ecotype Columbia-0 (Col-0), wordt gedurende de dag/nachtcyclus bestudeerd om de primaire metabolieten en hun functionele periodiciteit tijdens de circadiane cyclus te bepalen. Er is gebruik gemaakt van één- en twee-dimensionele HR-MAS methoden om de primaire metabolieten, waaronder organische zuren, suikers en aminozuren, te identificeren en

te kwantificeren. Multivariate analyse laat modules van primaire metabolieten zien die variëren gedurende de circadiane cyclus.

De HR-MAS NMR-workflow is toegepast in **hoofdstuk 3** om het metabole profiel op het midden van de dag vast te stellen voor de *Arabidopsis thaliana* mutanten VP16-02-003 en VP16-05-014 met verbeterde groeikenmerken. Er zijn achttien metabolieten geïdentificeerd die een significante verandering lieten zien van de concentraties tussen de mutanten en het wild-type Col-0. De biologische interpretatie van de metaboliet concentraties geeft aan dat de mutanten een verminderd vermogen hebben om te reageren op stress. Deze verminderde afweer leidt tot een aangepaste balans van groei en afweer in beide mutanten. Deze resultaten laten zien dat een hogere biomassa opbrengst kan worden behaald door de balans tussen groei en afweer aan te passen.

Hoofdstuk 4 beschrijft hoe de achttien biomarkers voor de verbeterde groeikenmerken van beide mutanten bestudeerd zijn tijdens de dag/nachtcyclus, omdat fotosynthese wordt geregeld door de circadiane klok. Het circadiane ritme van de metabolieten is nagenoeg hetzelfde voor de wild-type Col-0, de VP16-02-003 en de VP16-05-014 mutant en is dus robuust, terwijl de concentraties van de metabolieten verschillen tijdens de gehele circadiane cyclus. De biologische klok is niet afhankelijk van de balans tussen groei en afweer. Dit wijst erop dat het downstream fenotypisch aanpassen van bronnen en knelpunten, welke leiden tot groei, effectiever is dan het upstream aanpassen van de biologische klok.

Tenslotte bevat **hoofdstuk 5** een algemene discussie van dit werk wordt het toekomstperspectief beschreven. Samenvattend kan gesteld worden dat HR-MAS NMR-metabolomics gecombineerd met multivariate analyse een krachtig hulpmiddel is om het fenotype van planten te begrijpen en dat dit kan worden gebruikt voor *Arabidopsis* mutanten of andere gewassen die interessant zijn om te gebruiken als biologische bron voor bio-energie. In fluxomics kunnen metabolische fluxen, die het gevolg zijn van het samenspel van genexpressie, eiwitconcentratie, eiwitkinetiek en -regulatie, en metabolietconcentratie, gebruikt worden voor het interpreteren van complexe metabole routes en regulatie om een beter begrip te krijgen van het fenotype van planten. In combinatie met andere omics-technologieën, zoals transcriptomics en fluxomics, kan in de toekomst de complexiteit van een biologisch systeem verder worden opgehelderd.

Curriculum vitae

I was born the 2nd of January 1991 in Amsterdam. I did my secondary education at the Griftland College in Soest including biology, chemistry, physics, mathematics and informatics. In 2009, I started the bachelor Life Science & Technology at Leiden University and Delft University of Technology. During my bachelor thesis I came into contact with solid-state NMR for the first time, using it to study metabolic changes in the brain of an Alzheimer's disease mouse model under the supervision of Prof. dr. A. Alia. I discovered during this time, that I really love to do research so I decided to start the Research & Development master of Life Science & Technology at Leiden University in 2012. During my master's degree, I carried out two master research projects. The first one was at the Biophysical Structural Chemistry group of Leiden University under the supervision of Dr. N.S. Pannu working on X-ray crystallography. For my second master's thesis, I went to the European Bioinformatics Institute in Cambridge (UK). During this project, I worked on the transcriptomics data of T-helper cells in different stages of the cell cycle using bio-informatics tools under the supervision of Dr. S. Teichmann. I really like to combination of experimental work and bioinformatics.

This combination is what I found in my PhD research in the solid-state NMR group, which I started in Augustus 2014 under the supervision of Prof. dr. A. Alia and Prof. dr. H.J.M. de Groot on the FOM (now NWO-I) project "High-resolution solid-state NMR for understanding mechanism of high photosynthetic efficiency in engineered plants". During my PhD, I had the opportunity to present my work as posters at the BioSolar cell meeting (2015), Biophysics@ FOM (2015), NMR discussion meeting (2015, 2016), MetaboMeeting (2015, 2017) and the Reedijk Symposium (2017). For my poster at the NMR discussion meeting in 2015, I won the 3rd poster prize. Also, my poster at the MetaboMeeting 2017 was awarded a poster prize. I was also invited to present my work in an oral presentation at the BioSolar Cell meeting in 2015 about "Metabolic profiling of intact *Arabidopsis thaliana* leaves using ^{1}H high-resolution magic angle spinning NMR".

List of publications

D. Augustijn, U. Roy, R. van Schadewijk, H.J.M. de Groot, A. Alia. *Metabolic profiling of intact Arabidopsis thaliana leaves during circadian cycle using ^{1}H high-resolution magic angle spinning NMR*. PLoS One. 2016;11: e0163258

N. van Tol, M. Rolloos, **D. Augustijn**, A. Alia, H.J.M. de Groot, P.J.J. Hooykaas, B.J. van der Zaal. *An Arabidopsis mutant with high operating efficiency of Photosystem II and low chlorophyll fluorescence*. Sci Rep. Nature Publishing Group; 2017;7: 3314

N. van Tol, M. Rolloos, J.E. Pinas, C.V. Henkel, **D. Augustijn**, P.J.J. Hooykaas, B.J. van der Zaal. *Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors*. PLoS One. 2017;12: e0174236

D. Augustijn, N. van Tol, B.J. van der Zaal, H.J.M. de Groot, A. Alia. *Different growth-defence trade-off in Arabidopsis thaliana mutants with enhanced growth characteristics*. Submitted

D. Augustijn, N. van Tol, B.J. van der Zaal, H.J.M. de Groot, A. Alia. *Contrasting metabolite levels and a robust circadian rhythm in Arabidopsis thaliana mutants with enhanced growth characteristics*. Manuscript in preparation

D. Augustijn, N. van Tol, B.J. van der Zaal, H.J.M. de Groot, A. Alia. *Metabolic profiling of the Arabidopsis thaliana mutant LCF1*. Manuscript in preparation

D. Augustijn, H.J.M. de Groot, A. Alia. *Studying plant metabolomics by HR-MAS NMR*. Review manuscript in preparation

Acknowledgement

When I was choosing my bachelor degree, I knew quite quickly that I didn't want to study medicine and become a doctor. And now, years later, I will become a doctor by defending this PhD thesis. This is something I'm very proud of! But I didn't do it on my own and I would like to thanks some people which really help me through this PhD adventure.

First, I will start with my supervisors Prof. dr. H.J.M de Groot and Prof. dr. A. Alia. Thank you for supervising me, all the fruitful discussions and giving me directions but also the freedom to let it really become my project. I would like to thank Karthick and Fons for their help with the NMR experiments. I would like to thanks my fellow PhD researchers who made it a great time at the solid-state NMR group: Adriano, Emanuela, Fatemeh, Jan Paul, Maithili, Remco, Rubin, Vidya, Xinmeng, Yang and Zhongwu. And a special thanks to Jessica! Jessica, I enjoy our discussions about all kind of subjects and especially about women in science! You're an amazing person and there will come a day that women will change the world! Also, Liesbeth, thank you for being such an amazing person in this group. You were always so helpful! I would also like to thank Trudy. Your passion for the practical labs is very special. I really like that I can walk in and ask for something and you always have it. It makes the practical work so much easier.

During my bachelor's degree, I met Rianne and Femke. Together we always discuss amazing science articles and have passionate discussions about science. Thank you both for this and all the cups of tea we had the last years! Rianne, it was an honour to have an office next to you in the new building. All your help when I was stuck in writing and the possibility to talk to you once in a while were very helpful. You really helped me through this adventure. Another special woman is Marcella! You are always there, in good and in bad times. You're the friend everyone wishes they had. I also want to thank all the people of Chromosoom '09: Lisanne, Hannah, Paul, Alex, Frank & Chris! You're an amazing group of friends! Thank you for doing all things the last years to relax and not think about my PhD and just have fun. I also want to thanks all the people from scouting. Thank you for the possibility to develop myself in a non-scientific way and have fun with each other!

I also want to thank my family to stand behind me during my PhD research. I want to give special thanks my parents. They always believed in me and knew that I can do this! During my PhD, I lost my father, but I know for sure that you're proud of me! Toen mijn vader zijn master diploma haalde was hij er heel trots op dat hij niet meer titelloos wakker werd. Bij elk diploma wat ik behaalde, kreeg ik de volgende dag een berichtje hoe het was om niet meer titelloos wakker te worden. Nu ik mijn doctorstitel behaal, krijg ik geen berichtje van je. Maar ik weet zeker dat het geweldig is om wakker te worden als doctor!