NMR SPECTRAL ANALYSIS OF ASCORBATE-DEFICIENT MUTANTS OF *ARABIDOPSIS THALIANA*

Submitted by Felicia Charles Johnson
to the University of Exeter
as a dissertation for the degree of

Master of Philosophy

in Biological Sciences,

January, 2011.

This dissertation is available for Library use on the understanding that it is copyright material and that no quotation from the dissertation may be published without proper acknowledgement.

I certify that all material in this dissertation which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

........................
ABSTRACT

Ascorbate is an important antioxidant, involved in the scavenging of reactive oxygen species (ROS) such as hydrogen peroxide hydrogen. It is also a cofactor for 2-oxoglutarate-dependent dioxygenases, which are involved in the biosynthesis of a number of metabolites. In the current study, *Arabidopsis thaliana* ascorbate-deficient mutants (*vtc1*, *vtc2*, *vtc3*, and *vtc4*) were used to investigate the effect of ascorbate deficiency on the metabolic changes. NMR was used as an analytical technique for the metabolite profiling of *Arabidopsis thaliana vtc* mutants. A total of 45 NMR spectra representing wild type and four mutants of two different growth stages (rosette and flowering) were collected and spectral processed for exploratory analysis. Analysis of variance (ANOVA) and multivariate data analysis (hierarchical clustering) were used to find the differences between different strains and metabolite clustering patterns. Combinations of different parameters were used to find the best method for the analysis of NMR-based metabolites data. The method was developed which allowed making multiple comparison such as two growth stages and five strains. It was found the *vtc* mutants grouped separately from WT which contained the low ascorbate showing ascorbate deficiency might have some effect on the metabolic level. Two interesting resonances 1.32 and 1.33ppm were separated which were negatively correlated with the ascorbate levels in the *vtc* mutants. These resonances (1.32 and 1.33) are putatively identified as threonine and lactate. Further investigations are needed to investigate the nature of relationship between ascorbate and threonine and lactate.
# CONTENTS

## Chapter 1

### INTRODUCTION

- ASCORBATE METABOLISM AND FUNCTIONS 7
- ASCORBATE BIOSYNTHESIS (D-MAN/L-GAL PATHWAY) 9
- ASCORBATE BIOSYNTHESIS OTHER MULTIPLE PATHWAYS 10
- CONTROL OF ASCORBATE BIOSYNTHESIS 11
- ASCORBATE DEGRADATION 11
- ASCORBATE-DEFICIENT MUTANTS 11
- ISOLATION OF VTC MUTANTS 11
- ASCORBATE LEVELS IN VTC MUTANTS 12
- CHARACTERIZATION OF VTC MUTANTS 12
- PLANT METABOLOMICS 14
- METABOLITE/METABOLIC PROFILING 15
- METABOLOME TECHNOLOGIES FOR DATA ACQUISITION 16
- BIOINFORMATICS 21
- DATA ANALYSIS 21
- NMR SPECTRAL PROCESSING 22
- STATISTICAL ANALYSIS 23
- APPLICATION OF PLANT METABOLOMICS 24
- LIMITATIONS OF PLANT METABOLOMICS 26
- PRESENT STUDY 27
Chapter 2

MATERIALS AND METHODS

GERMINATION OF *ARABIDOPSIS THALIANA* SEEDS ON AGAROSE 29
TRANSFERRING *ARABIDOPSIS THALIANA* SEEDLINGS TO SOIL 32
HARVESTING OF *ARABIDOPSIS THALIANA* PLANTS FOR ANALYSIS 32
PREPARING *ARABIDOPSIS THALIANA* PLANTS FOR NMR ANALYSIS 33
ANALYSING *ARABIDOPSIS THALIANA* PLANTS BY NMR 34
NMR SPECTRAL PROCESSING WITH MESTRE-C 36
DATA ANALYSIS 37

Chapter 3

RESULTS AND DISCUSSION

NMR SPECTRAL FEATURES OF ASCORBATE- DEFICIENT MUTANTS 39
NMR DATA PRE-PROCESSING 41
MULTIVARIATE ANALYSIS OF ASCORBATE- DEFICIENT MUTANTS 42
DIFFERENTIALLY EXPRESSED BINS BETWEEN WT AND *VTC* MUTANTS 50

Chapter 4

CONCLUSION 54
REFERENCES 55
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1 Strategies for metabolomic analysis</td>
<td>16</td>
</tr>
<tr>
<td>Table 1.2 Common analytical techniques used in metabolomics</td>
<td>19</td>
</tr>
<tr>
<td>Table 2.1 Plant types used for the study</td>
<td>29</td>
</tr>
<tr>
<td>Table 2.2 Parameters for NMR Data Collection</td>
<td>35</td>
</tr>
<tr>
<td>Table 3.1 Clustering analyses of all plant types in rosette stage with 0.01 bin width and different processing combinations</td>
<td>45</td>
</tr>
<tr>
<td>Table 3.2 Clustering analyses of all plant types in rosette stage with 0.04 bin width and different processing combinations</td>
<td>46</td>
</tr>
<tr>
<td>Table 3.3 Clustering analyses of all plant types in flowering stage with 0.01 bin width and different processing combinations</td>
<td>47</td>
</tr>
<tr>
<td>Table 3.4 Clustering analyses of all plant types in flowering stage with 0.04 bin width and different processing combinations</td>
<td>48</td>
</tr>
<tr>
<td>Table 3.5 List of resonances showing discrimination in the spectral regions</td>
<td>51</td>
</tr>
<tr>
<td>Table 3.6 Resonances different between strains at spectral level with two bin widths</td>
<td>51</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1 The proposed pathways for L-ascorbic acid biosynthesis in plants</td>
<td>8</td>
</tr>
<tr>
<td>Figure 1.2 Wild type (Col-0) and four ascorbate-deficient mutants</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2.1 Spectral processing with MestRe-C</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.2 Schema of the data analysis</td>
<td>38</td>
</tr>
<tr>
<td>Figure 3.1 $^1$H-NMR spectra of the <em>Arabidopsis thaliana</em> (Wild type) extracted within D$_2$O:CD$_3$OD</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3.2a $^1$H-NMR spectra of <em>Arabidopsis thaliana</em> at rosette stage</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3.2b $^1$H-NMR spectra of <em>Arabidopsis thaliana</em> strains at flowering stage</td>
<td>41</td>
</tr>
<tr>
<td>Figure 3.3 HCA showing the difference in growth stages (rosette and flowering) and WT and the <em>vtc</em> mutants</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3.4 Heatmap for the $^1$H NMR based metabolic profiles of <em>Arabidopsis thaliana</em> Wild type and four ascorbate-deficient mutants in rosette</td>
<td>50</td>
</tr>
<tr>
<td>Figure 3.5 The intensity of the resonances 1.32 and 1.33ppm across the strains</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.6 Identification of the resonances 1.32 and 1.33 ppm in the original NMR spectra of one of the <em>vtc</em> mutant</td>
<td>53</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION

Ascorbate metabolism and functions

L-ascorbic acid (vitamin C) has a key role in prevention of scurvy in humans and also have shown to fulfil many other essential functions in animals and plants. The biological role of ascorbate (Asc) depends on its ability as a reducing agent, which acts as an effective antioxidant and a free radical scavenger (Smirnoff, 2011). It is the most abundant soluble antioxidant found in plants in almost all plant tissues, with the exception of dry seeds (De Gara et al., 1997) and is effective at directly scavenging reactive oxygen species (ROS) such as super oxide, hydrogen peroxide, singlet oxygen and the hydroxyl radical (Noctor & Foyer, 1998). It is a cofactor for various enzymes, including Fe^{2+}/2-oxoglutarate-dependent dioxygenases and violaxanthin de-epoxidase (De Tullio, 2004). Ascorbate is required for the regeneration of the lipophilic antioxidant α-tocopherol from the α -chromanoxyl radical (Fryer, 1992) and is also responsible for the regeneration of the carotenoid pigments: carotenes and xanthophylls (Eskling and Akerlund, 1997).
Significant progress has been made in the elucidation of the plant ascorbate biosynthetic pathway (Wheeler et al., 1998) to understand how ascorbate metabolism is controlled. These progresses provide a basis for engineering or manipulating its accumulation. The proposed pathway for L-ascorbate biosynthesis in plants is shown in Figure 1, D-Mannose/L-Galactose (D-Man/L-Gal) pathway is the predominant pathway. This pathway is different in animals where the last enzyme in the pathway (L-gulonolactone oxidase) is not expressed due to the lack of the enzyme. There are also other alternative biosynthetic pathways in plants along with the D-Man/L-Gal pathway and their existence in ascorbate biosynthesis is still under investigation. The alternative pathways are found to involve intermediates such as D-glucurono-1, 4-lactone, D-galacturonate, methyl-D-galacturonate and L-gulono-1,4-lactone (Smirnoff et al., 2001).
Figure 1.1 The proposed pathways for L-ascorbic acid biosynthesis in plants. The D-mannose/L-galactose pathway is shown in bold: D-mannose-1-phosphate is synthesised from D-glucose-1-phosphate via D-fructose-6-phosphate. It is converted into L-ascorbic acid via a series of steps catalysed by: 1, GDP-mannose pyrophosphorylase (Arabidopsis thaliana VTC1); 2, GDP-D-mannose-3, 5-epimerase; 3, GDP-L-galactose phosphorylase (Arabidopsis VTC2); 4, L-galactose-1-phosphate phosphatase (Arabidopsis thaliana VTC4); 5, L-galactose dehydrogenase; 6, L-galactono-1,4-lactone dehydrogenase. Alternative pathways and some of the enzymes catalysing the various reactions are shown: 7, GDP-D-mannose-4, 6-dehydratase; 8, GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase; 9, D-galacturonatereductase; 10, myo-inositol oxygcnase; 11, D-glucuronate reductase; 12, aldonolactonase; 13, L-gulonolactone dehydrogenase (Wheeler et al., 1998).
Ascorbate biosynthesis (D-Man/L-Gal pathway)

Ascorbate is present in mill molar (mM) concentration in plant tissues which varies considerably between tissues. The concentration of ascorbate also depends on the physiological status of the plant as well and on the environmental factors (Luwe et al., 1993). Advances in understanding of biosynthetic pathway of Asc in plants have been made by studying Pisum sativum and Arabidopsis plants (Wheeler et al., 1998). Plants synthesize ascorbate involving D-glucose-6-P, D-fructose-6-P, D-mannose-1-P, GDP-D-mannose, GDP-L-gulose, GDP-L-galactose, L-galactose-1-P and L-galactose as intermediaries in the production of L-galactono-1,4-lactone, a precursor, oxidised to ascorbate by the enzyme L-galactose 1,4 lactone dehydrogenase. There are many enzymes involved which catalyses various steps in Asc biosynthesis. Some of those key enzymes are briefly described below. Phosphomannose isomerase (PMI), is involved in the conversion of fructose 6-P to mannose 6-P and its activity is readily detected in Arabidopsis leaf extracts and also in other species (Smirnoff and Wheeler, 2000). Phosphomannose mutase (PMM) catalyses the interconversion of mannose 6-P and Mannose 1-P. Studies show increase in ascorbate when PMM was over expressed in Nicotiana benthamiana, Arabidopsis and virus-induced gene silencing caused a decrease in ascorbate (Smirnoff, 2011). GDP-mannose pyrophosphorylase (GMP) catalyses the formation of GDP-mannose 1-P and GTP, GDP-mannose-3, 4-epimerase (GME) catalyses the double epimerisation of GDP-mannose with the production of GDP-L-galactose.
GDP-L-galactose phosphorylase/guanylyl transferase was identified in pea seedling extracts (Dowdle et al., 2007; Ishikawa et al., 2006a and b) and in Arabidopsis by (Dowdle et al., 2007; Laing et al., 2007 and Linster et al., 2007, 2008) catalyse the phosphorylytic breakdown of GDP-L-galactose to L-Galactose 1-P, L-galactose 1-P phosphatase, hydrolyses L-galactose 1-P to L-Gal. L-galactose dehydrogenase (L-GalDH) catalyses NAD\(^+\)-dependent oxidation of L-Gal at C1 to produce L-galactono lactone (Gatzek et al., 2002) and L-galactono-1,4 lactone dehydrogenase (L-GalLDH) catalyses the last step which is the oxidation of L-GalL to ascorbate.

**Ascorbate biosynthesis Other Multiple Pathways**

It has been proposed that plants not only synthesize ascorbate from D-Man/L-Gal but also from D-galacturonic acid (D-GalUA), L-gulose and from myo-inositol/D-glucuronic acid. External application of uronic acid derivatives, labelling studies, gene expression studies, over expression studies have shown to increase in ascorbate providing evidence for the existence of other pathways involved in ascorbate biosynthesis (Smirnoff, 2011).

**Control of Ascorbate biosynthesis**

The rate of biosynthesis depends on the amount of each enzymes and kinetic properties in relation to substrate concentrations and other factors. Ascorbate concentration is regulated by cell type and environmental conditions. Ascorbate concentration is high in leaves, peel of fruits (Davey et al., 2000; Li et al., 2008), meristem cells (Cordoba-Pedregosa et al., 2003) germinating seeds (Pallanca and Smirnoff, 2000) and also in high light (Gautier et al., 2009;
Li et al., 2009; Dowdle et al., 2007), low temperature (Schoner and Krause, 1990), jasmonates Wolucka et al., 2005) mechanical wounding (Suza et al., 2010) and also when over expressing biosynthesis genes (Smirnoff, 2011) however, a very little is known about how the biosynthesis or breakdown is controlled.

**Ascorbate degradation**

Asc is generally involved as an electron donor transferring electrons by oxidation as monodehydroascorbate (MDHA) and dehydroascorbate (DHA). MDHA does not readily react with other molecules and is relatively harmless. DHA undergoes spontaneous and irreversible hydrolysis to 2,3-diketogulonic acid (Washko et al., 1992). Ascorbate and DHA are catabolised in plants and give rise to a number of end products, including L-threonate, oxalate and L-tartrate (Loewus, 1999).

**Ascorbate-deficient mutants**

To get a better understanding of the biosynthetic pathways mutants have been a valuable tool and are utilized to uncover a number of plant primary and secondary metabolic pathways and to identify the pathways of ascorbate catabolism. Gene function can be discovered through the use of mutants that are altered in gene of interest and the physiological and biochemical changes are observed comparing to the Wild type. Conklin et al., (2000) identified a number of ascorbate-deficient mutants (Figure 2a, 2b).
Isolation of \textit{vtc} mutants

The \textit{vtc} (vitamin-c) mutants were produced through ethyl methane sulfonate mutagenesis of the Col-0 wt seeds, which is the Wild type plants of Arabidopsis and the plants were screened by nitro blue tetrazolium (NBT) based assay for an Asc-deficient phenotype and the produced mutants were also used for the analyses of genetic segregation and allelism. Asc deficiency in the mutants \textit{vtc}1-2, \textit{vtc}2-1, \textit{vtc}2-2, \textit{vtc}3-1 and \textit{vtc}4-1 is conferred by single monogenic recessive traits. \textit{vtc} mutants represent five different loci; \textit{VTC}1, \textit{VTC}2, \textit{VTC}3, \textit{VTC}4 AND \textit{VTC}5. Two of the mutants (\textit{vtc}1-1 and \textit{vtc}2-1) were isolated through their increased ozone sensitivity. In total there are two \textit{vtc}1 mutant alleles, three \textit{vtc}2 mutants as well as \textit{vtc}3-1 and \textit{vtc}4-1. \textit{vtc}5-1 and \textit{vtc}5-2 were produced by T-DNA insertional mutagenesis (Linster and Clarke, 2008).

The \textit{VTC}1 locus has been cloned and it encodes GDP-mannose pyrophosphorylase (GMP) (Conklin \textit{et al.}, 1999) which is an enzyme that catalyses the conversion of D-mannose-1-phosphate to GDP-mannose in the ascorbate biosynthesis pathway. The \textit{vtc}1 mutant has a decreased ability to convert glucose and mannose to ascorbate.

\textit{VTC}2, \textit{VTC}5 and \textit{VTC}4 have been identified as genes of the D-Mannose/L-Galactose ascorbate biosynthesis pathway, encoding respectively GDP-L-galactose phosphorylase; GDP-β-L-Gal: orthophosphate guanyltransferase (Dowdle \textit{et al.}, 2007; Ishikawa \textit{et al.}, 2006a and b) and L-galactose 1-P phosphatase (Conklin, \textit{et al.}, 2006).
Ascorbate levels in *vtc* mutants

The *vtc* mutants contain 1/3 to 1/2 the total Asc present in the wild type Col-0 (Conklin et al., 2000). They range from 50-70 % of wild-type levels in *vtc2-3*, *vtc3* and *vtc4*, to 25-30% in *vtc1-1*, *vtc1-2* and 10-20% in *vtc2-1* and *vtc2-2*. The Asc levels increases upon the transition from vegetative to reproductive state (Gander 1982). The reproductive tissues contain twice the amount of Asc found in mature leaves of both wt and mutants.

Characterization of *vtc* mutants

There has been many studies that provides strong evidence that low ascorbate reduces growth and increases susceptibility to a range of stresses. Characterization of *vtc1-1* and *vtc2-1* has revealed that both the mutants have slow growth rate (Pavet *et al.*, 2005; Conklin *et al.*, 2000). *vtc1-1* and *vtc2-1* have been widely studied in relation to the stress responses, while there is little information on other mutants. The Wild type Arabidopsis is quite tolerant to ozone (O₃), probably because these plants mount an effective antioxidant response (Sharma and Davis, 1994; Conklin and Last, 1995). *vtc1-1* mutant is extremely O₃ sensitive with visible injury including lesion formation, enhanced chlorosis, and/or tissue collapse. *vtc2* mutants have varied O₃ sensitive phenotypes. *vtc2-1* was similar to *vtc1-1*, *vtc2-3* appeared somewhat sensitive but *vtc2-2* was not visibly injured (Conklin *et al.*, 2000). All the four mutants are more salt-sensitive (Huang *et al.*, 2005; Smirnoff, 2000). Both *vtc*1 and *vtc*2 has reduced basal thermo tolerance (Larkindale *et al.*, 2005). *vtc2-1* has greater thermo induced photo emission indicating increased lipid peroxidation at high temperature (Havaux, 2003). In *vtc* 1-1, 171 genes were found differentially expressed and identified by transcriptomic analysis of which 12.9% genes were involved in cell defence (Pastori, 2003).
vtc1-1 and vtc2-1 have increased resistance to infection by virulent pathogens (Barth et al., 2004; Pavet et al., 2005). Ascorbate deficient mutants also show increased salicylic acid, increased expression of pathogenesis-related proteins, peroxidise activity and accumulation of the phytoalexin camalexin (Kleibenstein, 2004).

**Figure 1.2 Wild type (Col-0) and four ascorbate-deficient mutants (vtc 1, vtc 2, vtc 3 and vtc 4) (vtc= vitamin c mutants).**

**Plant metabolomics**

Metabolic networks are very complex as there is an enormous chemical diversity of compounds. In the plant kingdom there are over approximately 200,000 metabolites and about 5000 individual low molecular weight compounds (Fiehn, 2002). Metabolomics is the "systematic study of the unique chemical fingerprints that specific cellular processes leave behind" - specifically, the study of their small-molecule metabolite profiles (Bennett, 2005). Metabolomics has emerged as an important functional genomics tool to aid in the comprehensive analysis of all the metabolites in the cellular system adding to the understanding of the complex molecular interactions in the biological systems (Sumner et al., 2003; van der Greef et al., 2003; Saito and Matsuda, 2010). Metabolomics currently in
plants are more complex due to large number of metabolites and the presence of
differentiated tissues, including specialist storage organs, with different metabolite
compliments (Ward et al., 2002). The field of plant metabolomics is a complicated
interdisciplinary research field that requires bioscience, analytical chemistry, organic
chemistry, chemometrics, and informatics knowledge. Genome and transcript or protein
profiling offers only limited information about the chosen system but profiling metabolome
actually provides the most “functional” information (Sumner et al., 2003). The ultimate aim
of metabolomics is an unbiased and comprehensive monitoring of all the metabolites in a
cellular system and to nearly complete the molecular picture of the state of a particular
biological system at a given time (Fiehn 2002). There appear to be differences of opinion as
how best to define the comprehensive profiling of the metabolome. Sumner (2003) proposed
that any technology whose output is processed with pattern recognition software and without
differentiation of individual metabolites should be termed metabolic finger printing not
metabolomics or metabonomics. Metabolic finger printing which is also the chemometric
approach where the compounds are not initially identified only their spectral patterns and
intensities are recorded, statistically compared and used to identify the relevant spectral
features that distinguish sample classes (Wishart, 2008; Xia et al., 2009)
Metabolite/metabolic profiling

Metabolite/metabolic profiling is the measurement of hundreds or potentially thousands of metabolites in other words a quantitative approach which aims to formally identify and quantify all detectable metabolites from the spectra, that requires the compounds of interest to be known from a spectral reference library obtained from authentic standards prior to subsequent data analysis (Xia et al., 2009). There is yet another approach of profiling metabolome called targeted analysis which aims to measure the concentration of a limited number of known metabolites precisely after knowing the structure of the target metabolite which also requires the compounds of interest to be known a priori in purified form (Shulaev, 2006). The strategies for metabolomic analysis given by Dunn and Ellis (2005) is presented in Table 1.1 below.
### Table 1.1 Strategies for metabolomic analysis (Dunn and Ellis, 2005)

<table>
<thead>
<tr>
<th>Metabolomics</th>
<th>Non-biased identification and quantification of all metabolites in a biological system. Sample preparation must not exclude metabolites, and selectivity and sensitivity of the analytical technique must be high.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolite profiling</td>
<td>Identification and quantification of a selected number of pre-defined metabolites, generally related to a specific metabolic pathway(s). Sample preparation and instrumentation are employed so to isolate those compounds of interest from possible matrix effects prior to detection, normally with chromatographic separation prior to detection with MS. Widely used in pharmaceutical industry.</td>
</tr>
<tr>
<td>Metabolic fingerprinting</td>
<td>High-throughput, rapid, global analysis of samples to provide sample classification. Quantification and metabolic identification are generally not employed. A screening tool to discriminate between samples of different biological status or origin. Sample preparation is simple and, as chromatographic separation is absent, rapid analysis times are small (normally 1 min or less)</td>
</tr>
<tr>
<td>Metabolite target analysis</td>
<td>Qualitative and quantitative analysis of one or a few metabolites related to a specific metabolic reaction. Extensive sample preparation and separation from other metabolites is required and this approach is especially employed when low limits of detection are required. Generally, chromatographic separation is used followed by sensitive MS or UV detection</td>
</tr>
<tr>
<td>Metabonomics</td>
<td>Evaluation of tissues and biological fluids for changes in endogenous metabolite levels that result from disease or therapeutic treatments</td>
</tr>
</tbody>
</table>


Metabolome technologies for data acquisition

The central to metabolomic processes are a variety of chemical profiling technologies. The gross chemical composition of various biological fluids has been investigated by a variety of chromatographic (separation methods) and spectroscopic techniques (detection methods), notably gas and liquid chromatography (Fiehn et al., 2000b; Sauter et al., 1991; Roessner et al., 2001), Mass spectrometry (Matsumoto and Kuhara, 1996; Wolfender and Hostettmann, 1996; Aharoni et al., 2002), NMR (Ward et al., 2002, Kikuchi et al., 2004) and Infrared spectrophotometry (Jackson and Mantsch, 1996). Various technologies used are discussed and their advantages and disadvantages (Shulaev, 2006) are shown in Table 2 below:

Mass spectrometry (MS): Due to its high sensitivity and wide range of covered metabolites, MS has become the technique of choice in many metabolomics studies (Halket et al., 2005). MS can be used to analyse biological samples either directly via direct-injection MS or following chromatographic or electrophoretic separation. Direct-injection MS, especially when using a high-resolution mass spectrometer, provides a very rapid technique to analyse large number of metabolites, and therefore is extensively used for metabolic fingerprinting and metabolite profiling. Mass spectrometry has a serious drawback, known as ionization suppression (which results from the presence of less volatile compounds that can change the efficiency of droplet formation or droplet evaporation, which in turn affects the amount of charged ion in the gas phase that ultimately reaches the detector) which might diminish quantitative reproducibility. To avoid this problem and to decrease the complexity of the sample, MS is often used as a hyphenated technique, i.e. it is coupled with gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE) where
the sample mixture is separated by chromatography or electrophoresis, and individual compounds or less complex mixtures of compounds (which cannot be separated due to similar properties) that are eluted from the column or capillary are analysed by MS.

**Gas chromatography**: is one of the most widely used and is a powerful method especially when interfaced with electron impact (EI) quadruple or time-of flight (TOF) mass spectrometry (GC-MS). GC-MS is viewed as the gold standard of metabolomic techniques due to the high separation power of GC combined with better deconvolution. The advantage of this coupled approach is that, it is possible to profile several hundred chemically diverse compounds including organic acids, most amino acids, sugars, sugar alcohols, aromatic amines and fatty acids (Roessner, 2000). It offers very high chromatographic resolution, but requires chemical derivatization (chemical conversion of sample compounds into their derivatives which makes them suitable for separation and analysis) for many bio molecules and hence some large and polar metabolites cannot be analysed by GC (Schauer *et al.*, 2005).

**Liquid chromatography/mass spectrometry (LC-MS)**: LC-MS is being increasingly used due to its high sensitivity and a range in analyte polarity and molecular mass wider than GC-MS. Liquid chromatography, and specifically high-performance liquid chromatography (HPLC), combines both high resolution and analytical flexibility. It can be tailored for the analysis of a specific metabolite or class of compounds, or it can be used for the analysis of a broad range of compound classes. LC-MS has one advantage over GC-MS, in that; there is largely no need for chemical derivatization of metabolites (Shulaev, 2006).
Capillary electrophoresis-mass spectrometry (CE-MS): CE-MS provides several important advantages over other separation techniques. It has a very high resolving power; very small sample is required for analysis and has short analysis time. CE has been used for both targeted and non-targeted analysis of metabolites. One of the significant advantages of the CE-MS is the ability to separate cations, anions and uncharged molecules in a single analytical run, and therefore can be used for simultaneous profiling of many different metabolite classes (Soga *et al*., 2002). This feature makes it a very attractive and promising analytical technique for high-throughput non-targeted metabolomics.

Fourier transform infrared spectroscopy (FTIR): is useful in the field of large-scale exhaustive profiling as an easy and high-throughput method (Gidman *et al*., 2003; Johnson *et al*., 2003). It is used for the quantification of compounds that have specific functional groups although it is not optimal for complicated metabolite profiling due to the lack of a separation procedure. Fourier transformation ion cyclotron resonance mass spectrometry (FT-ICR-MS), is the latest mass spectrometry technology. Infusion FT-ICR-MS analysis without pre-separation by chromatography has been achieved for exhaustive metabolic profiling (Aharoni, *et al*., 2002).
<table>
<thead>
<tr>
<th>Analytical method</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>Rapid analysis</td>
<td>Low sensitivity</td>
</tr>
<tr>
<td></td>
<td>High resolution</td>
<td>Convoluted spectra</td>
</tr>
<tr>
<td></td>
<td>No derivatization needed</td>
<td>More than one peak per component</td>
</tr>
<tr>
<td></td>
<td>Non-destructive</td>
<td>Libraries of limited use due to complex matrix</td>
</tr>
<tr>
<td>GC/MS</td>
<td>Sensitive</td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>Robust</td>
<td>Often requires derivatization</td>
</tr>
<tr>
<td></td>
<td>Large linear range</td>
<td>Many analytes thermally-unstable or too large for analysis</td>
</tr>
<tr>
<td></td>
<td>Large commercial and public libraries</td>
<td></td>
</tr>
<tr>
<td>LC/MS</td>
<td>No derivatization required (usually)</td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>Many modes of separation available</td>
<td>Limited commercial libraries</td>
</tr>
<tr>
<td></td>
<td>Large sample capacity</td>
<td></td>
</tr>
<tr>
<td>CE/MS</td>
<td>High separation power</td>
<td>Limited commercial libraries</td>
</tr>
<tr>
<td></td>
<td>Small sample requirements</td>
<td>Poor retention time reproducibility</td>
</tr>
<tr>
<td></td>
<td>Rapid analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can analyse neutrals, anions and cations in single run</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No derivatization (usually)</td>
<td></td>
</tr>
<tr>
<td>FTIR</td>
<td>Rapid analysis</td>
<td>Extremely convoluted spectra</td>
</tr>
<tr>
<td></td>
<td>Complete fingerprint of sample chemical composition</td>
<td>More than one peak per component</td>
</tr>
<tr>
<td></td>
<td>Metabolite identification nearly impossible</td>
<td>No derivatization needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requires sample drying</td>
</tr>
</tbody>
</table>
Nuclear magnetic resonance (NMR): can uniquely identify and simultaneously quantify a wide range of organic compounds in the micromolar range (Viant et al., 2003; LeGall et al., 2003; Krishnan et al., 2005). In recent times, high-resolution NMR well-suited to metabolomics studies has become a key technology for the elucidation of biosynthetic pathways and metabolite flux via quantitative assessment of multiple isotopologues (Eisenreich and Bacher, 2007). NMR spectroscopy is the only detection technique which does not rely on separation of the analytes, since the sample preparation for NMR is straightforward and also is non-destructive the sample can thus be recovered for further analyses (Bailey et al., 2003). All kinds of small molecule metabolites can be measured simultaneously - in this sense, NMR is close to being a universal detector.

$^1$HNMR is also a rapid technique that characterizes and quantifies a broad range of metabolites in a non-targeted way, since in principle any chemical species that contains protons gives rise to signals. NMR signals (i.e. resonance) are observed when a sample is irradiated with pulses of radiofrequency electromagnetic radiation in a strong magnetic field. Each nucleus within a molecule experiences a slightly different magnetic field because of its distinct chemical environment and absorbs energy at a slightly different frequency. The separation of these resonance frequencies from an arbitrarily chosen reference is called the chemical shift (ppm). The result of an NMR analysis of a tissue is a spectrum, that is, a plot of intensity (area) against chemical shift in which each signal occurs at a characteristic energy. These signals are profiled as NMR spectrum comprises peaks which, for a given solvent appear at characteristic frequencies. The major limitation of NMR for comprehensive metabolite profiling is its relatively low sensitivity, making it inappropriate for the analysis of large number of low-abundance metabolites (Sumner et al., 2003). Some of the studies using NMR has shown to cover most of the “organic” compounds such as
carbohydrates, amino acids, organic and fatty acids, amines, esters, ethers and lipids, which are present in plant tissues (Ward et al., 2002) and when coupled with statistical analysis serves as an excellent screen to analyse metabolome in a non-destructive way. NMR has also provided information on biosynthesis (Lutterbach and Stockigt, 1995; Prabhu et al., 1996; Fiehn and Weckwerth, 2002), on metabolism (Ratcliffe and Shachar-Hilt, 2001), and on the effects of herbicides on metabolism (Lutterbach and Stockigt, 1994 and 1995) and mode-of-action (Hole et al., 2000, Hadfield et al., 2001), or used in investigations of whole plants (Schneider, 1997; Pope et al., 1993). Although NMR and MS are most often used for large-scale analysis, metabolomics is not limited to these techniques. Other alternatives include thin-layer chromatography (Tweeddale et al., 1998 and 1999). HPLC with UV/visible absorbance, photodiode array (PDA) (Fraser et al., 2000) or electrochemical detectors (Gamache et al., 2004, Kristal et al., 2002) FT-IR (Gidman et al., 2003), phenotype microarrays (Bochner et al., 2001) and variety of other enzymatic assays. Combined use of multiple techniques or multiple detectors in online or parallel analysis can significantly increase the metabolite coverage, increase quantification limits and improve identification of metabolites from a single biological sample.
**Bioinformatics**

**Data Analysis**

The major challenge of metabolomics studies is data processing and data interpretation (Shulaev, 2006). Database management systems for metabolomics are required to collect both metadata (metadata is the data about the experiment), raw and processed experimental data. Storing metadata, covering experimental design, the nature of the samples and their treatment prior to the analysis and information about the analytical technique and data-processing details are important to be able to reproduce the experimental conditions and compare results obtained in different laboratories (Bino et al., 2004). All the metabolomic approaches heavily rely on bioinformatics for the storage, retrieval, and analysis of larger datasets and many efforts have been made on software development (Katajamaa and Oresic, 2007). There are many tools used to align, visualize and differentiate components in datasets and the individual components are then correlated and placed in metabolic networks or pathways. Changes in metabolite levels may be drastic or subtle and those subtle changes will require statistical processing to determine whether the observed changes are significant or not (Miller and Miller, 2000).

For NMR-based metabolomics approach, after NMR experimental data collection, post experimental data handling including NMR spectral processing, data pre-processing and data analysis, is critical for obtaining good results. To assist these procedures, several commercial software tools have been released, such as: AMIX (Bruker Biospin, Germany), KnowItALL (BIO-RAD, USA), Chenomx NMR Suite (Chenomx, Canada) and Hi-Res (Zhao, 2006). NMRPipe is a widely used traditional NMR data processing software tool (Delaglio et al., 1995). These software tools provide plenty of functions involving spectral
processing, comprehensive identification and quantification of metabolites. However, due to the complexity of NMR data and different application purposes, further data analysis procedures, such as filtering out unwanted variations (e.g. background noise, uncorrelated variation in data model, etc.) in a dataset, generating and applying predictive classification or regression models, are usually required. To complete these tasks, researchers usually invoke other advanced chemometrics tools or statistical tools. Commercial software packages such as MATLAB (Mathworks, USA), SIMCA-P (Umetrics, Sweden) and SPSS (SPSS, USA) are frequently used by researchers.

**NMR spectral processing**
Spectral data often include some disturbance according to each specificity and environmental perturbation. Such disturbance needs to be corrected for by appropriate data pre-processing. Several tactics, such as spectral difference, noise reduction, baseline correction, normalization, mean centre, and scaling are often performed in common pre-processing as for the metabolomics studies, spectral processing may have significant impact on data analysis (Deferez and Colquhoun, 2003). The present study uses a commercial software MestRe-C (Magnetic Resonance Companion) Cobas and Sardina, 2003) designed specifically for metabolomics, will be evaluated.

**Statistical Analysis**
Statistical analyses must be performed to ensure good analytical rigor, but unfortunately are often a burden when working with large datasets. Tools are therefore used for high throughput statistical analyses of all components in a dataset to provide sound evidence for the relevance of changes in a metabolite level contained in the raw data. Computer based applications are used that can differentiate whether or not samples are statistically similar or different and what the exact differences/similarities are. A single metabolite profile can yield
hundreds of distinct components. This wealth of information that needs to be interpreted, leads to significant challenges in processing the data. To simplify the task, many researchers have used techniques to reduce the dimensionality of the data set and to visualize the data (Sumner et al., 2003).

Principle Component Analysis (PCA) is the choice for metabolomics for many researchers as the data is complex (Le Gall, 2001; Ward et al., 2002; Defernez and Colquhoun, 2003). PCA provides a way to summarize the information contained in large sets of spectra. It transforms the initial variables (or ‘measurements’) into a much smaller set of variables or PC scores. The main purpose of PCA is to eliminate the collinear problem and then reduce the dimensionality of the original feature (variable) space. It is an unsupervised method used to reveal the internal structure of datasets in an unbiased way. PLS is a supervised method for regression. PCA and PLS reveal the variable contribution for the separation between different groups by loadings or regression coefficients (Fukusaki and Kobayashi, 2005). Machine learning methods such as self-organising maps (SOM) and artificial neural networks (ANN) are also promising in research (Fukusaki and Kobayashi, 2005). A variety of approaches have been applied for the statistical analysis of spectral data. One approach is to gather information (e.g. integrate individual peak areas) for as many constituents as possible and study the occurrence of each of these compounds separately, for example by calculating coefficients of variation or carrying out an Analysis of variance (ANOVA) (Le Gall, 2001; Smilde et al., 2005). ANOVA estimates the factor effects and tests significance by splitting the variations in orthogonal and independent parts (Searle, 1971). Another data analysis approach, Hierarchical cluster analysis (HCA), is a method of cluster analysis based on the multivariate distance between every pair of data points. In HCA, the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place,
which may run from a single cluster containing all objects to n clusters each, containing a single object (Mochida et al., 2009). For the present study, the statistical analysis was carried out and visualized in an application tool called Multi Experiment Viewer (MeV) (Saeed et al., 2006).

In plant metabolomics, bioinformatics tools provide methods of determining differences or similarities in datasets. The metabolomic data sets must then be integrated and correlated in a global manner with genetic and enzymatic data, pathways assembled into systems, and literature references incorporated as learning tools to annotate existing data to yield in silico biological information (Palsson, 2000).

**Application of Plant metabolomics**

Although relatively young and still very much in development, plant metabolomics is now being widely applied and is already considered as a technology, which is a ‘maturing science’ and is ‘established and robust’ (Dixon et al., 2006; Schauer and Fernie, 2006). Some of the general applications of plant metabolomics are studied in plant tissues using NMR (Kim et al., 2010; Kruger et al., 2008) GC-MS (Lisec et al., 2006), LC-MS (De Vos et al., 2007). Hall et al., (2008), has elaborated the use of plant metabolomics and its potential application for human nutrition. Metabolite profiling has been performed on a diverse array of plant species, including tomato (Schauer et al., 2005; Tunali et al., 2003), potato (Roessner et al., 2000, Roessner et al., 2001), rice (Sato et al., 2004), wheat (Hamzehzarghani et al., 2005), strawberry (Aharoni, et al., 2002), medicago (Chen et al., 2003), cucumber (Tagashira et al., 2005), lettuce (Garratt et al., 2005), tobacco (Blount et al., 2002), poplar (Robinson et al., 2005, Brosche et al., 2005) and eucalyptus (Merchant et al., 2006). Metabolite profiling has also been carried out widely in Arabidopsis thaliana, in a tissue specific level (Schad et
al., 2005), in transgenic plants (Le Gall et al., 2005) and also for comparison of ecotypes of Arabidopsis thaliana (Ward et al., 2002). Hirai et al., (2004) linked the genomic data and the function of metabolite under the deficiency of sulphur and nitrogen in Arabidopsis thaliana.

Metabolomics has been useful to ascertain, for example, the metabolic response to herbicide (Ott et al., 2003), the equivalence of genetically modified (GM) and conventional crops (Catchpole et al., 2005; Defernez et al., 2004; Barros et al., 2010; Matsuda et al., 2010) and the classification of plant genotypes (Roessner et al., 2000; Roessner et al., 2001, Fiehn et al., 2000a, Tikunov, et al., 2005, Kim, et al., 2010), Quality control of medicinal plants viz., Gingko (Agnolet et al., 2010), St. John’s wort (Roos et al., 2004), Rasmussen et al., 2006), Ginseng (Yang et al., 2007), and to find activity-related compounds in medicinal plants (Cardoso-Taketa et al., 2008, Modarai et al., 2010) and to study the plant interaction with other organisms (Jahangir et al., 2008, Abdel-Farid et al., 2009, Leiss et al., 2009). Also there have been significant recent successes in gaining new insights into the pathways of lignin biosynthesis (Anterola and Lewis, 2002) and isoflavone synthesis (Jung et al., 2000; Liu et al., 2002) and progression in metabolite-profiling methods for isoprenoids (Lange et al., 2001), oxylipins (Weichert et al., 2002), taxane diterpenoids (Ketchum et al., 2003), alkaloids (Yamazaki, 2003), flavonoid glycosides (Le Gall et al., 2003) and volatile compounds (Verdonk et al., 2003; Flamini et al., 2003; Brown, 2002).

More recently, metabolomics has been much used in descriptions of the response of plants to a wide range of biotic or abiotic stresses, to decipher gene function, to investigate metabolic regulation and (in combination with analysis of other molecular entities of the cell) as part of integrative analyses of the systemic response to environmental or genetic perturbations (Schauer and Fernie, 2006).
Limitations of plant metabolomics

There are several critical factors in metabolomics studies, since the metabolome is sensitive to perturbation. Plant cultivation, sampling, extraction and derivatization must all be carefully standardised and strategies need to be incorporated to minimize variations (Fukusaki and Kobayashi, 2005). It is generally accepted that a single analytical technique will not provide sufficient visualization of the metabolome and therefore, multiple technologies are needed for a comprehensive view and sometimes a combination of techniques are to used (Le Gall et al., 2005). The selection of the most suitable technology is generally a compromise between speed, selectivity and sensitivity (Sumner et al., 2003). Metabolic profiling is complex and it involves various techniques to cover a wide range of metabolites. Each metabolite should be considered according to its characteristics in the following categories: hydrophilic, hydrophobic, small molecule, large molecule, charged, uncharged and combinations of these (Schauer and Fernie, 2006). Metabolites vary due to volatility, polarity, solubility and chromatographic behaviour mean that multiple methods will need to be deployed to analyse different subsets of metabolites (Ward et al., 2002). A major technological challenge encountered in metabolomics is dynamic range. Dynamic range defines the concentration boundaries of an analytical determination over which the instrumental response as a function of analyte concentration is linear. The dynamic range of many techniques can be severely limited by the sample matrix or the presence of interfering and competing compounds. In other words, the presence of some excessive metabolites can cause significant or severe chemical interferences that limit the range in which other metabolites may be successfully profiled. For example, high levels of primary metabolites such as sugars often interfere with the ability to profile secondary
metabolites such as flavonoids. This is one of the most difficult issues to address in metabolomics (Sumner et al., 2003). There is another drawback where metabolomics is a snapshot and thus is not particularly suited to studying the rapid kinetics of metabolic processes and requires measurement of fluxes (a series of snapshots), which represents a major challenge (Kim et al., 2011). A truly comprehensive analysis of the metabolome is currently not feasible and the number of the primary and secondary metabolites in any given plant species is still uncertain. This is the current inability and limitation of metabolomics. Metabolomics is not a goal in itself; it is a tool for improving our understanding of the metabolism and biochemistry of organisms. To provide a deeper biological meaning, metabolomic studies should deal with identified metabolites instead of unknown signals: the number of signals is less important than the number of identified metabolites. To facilitate identification, an important issue and a great challenge will be to build a common public database to share information within scientific communities (Kim et al., 2011)

**Present study**

The recent progress about the pathways, control of ascorbate metabolism and the potential roles of intracellular and long distance transport as factors determining ascorbate concentration has helped in understanding plant ascorbate to a certain extent. There are also certain areas of ascorbate metabolism which require a deeper understanding. Much remains to be learned about the contribution of alternative ascorbate biosynthetic pathways in each plant tissue and the regulation of metabolic fluxes, and many important genes have not yet been identified. There are many important questions that still needs to be explored about ascorbate metabolism and the ascorbate-deficient mutants provide an opportunity to
investigate the following questions such as: What is the effect of each specific mutation on metabolism and can the function of VTC3 identified? Ascorbate has proposed roles as an antioxidant and as a cofactor for dioxygenase enzymes that are involved in synthesis of a variety of plant hormones and secondary compounds (e.g. phenolic compounds involved in defence against pathogens). Therefore can the effect of ascorbate deficiency on plant function, as visualised by the metabolome, provide information on the function(s) of ascorbate? Ascorbate mutants are more sensitive to environmental stresses and, if so, can this be detected by changes in their metabolome compared to wild-type plants. Profiling of all the metabolites contained in the ascorbate deficient mutants will offer much useful information about metabolism and manipulation. The main objectives of the present study are: 1. to use NMR as an analytical tool for profiling the vtc mutants and Wild type plants to produce NMR data 2. to use software packages MestReC to perform spectral processing to study the ascorbate NMR data 3. to look for differences in spectral processed data at the metabolite levels across the plant types (Wild type/ mutants and alleles), growth stages (rosette and flowering) 4. to verify the discriminating bins from the statistical analysis with the original spectra for the presence of peaks.
Chapter 2

MATERIALS AND METHODS

*Arabidopsis thaliana* ecotype Columbia-0 seeds (Wild type) and ascorbate-deficient mutants namely vtc1, vtc2, vtc3 and vtc4 were chosen for the present study (Table 2.1). All the mutants have the same genetic background.

<table>
<thead>
<tr>
<th>Wild type plant</th>
<th>Ascorbate-deficient Mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>vtc1</td>
</tr>
<tr>
<td></td>
<td>vtc2</td>
</tr>
<tr>
<td></td>
<td>vtc3</td>
</tr>
<tr>
<td></td>
<td>vtc4</td>
</tr>
</tbody>
</table>

The seeds were grown, seedlings transplanted to soil, harvested, prepared and extracted for NMR analysis in The National Centre for Plant and Microbial Metabolomics, at Rothamsted Research. U.K. The protocols followed for each of the above stages are given below:

**Germination of Arabidopsis thaliana seeds on agarose:**

Before starting the procedure Laminar Flow Hood was turned on and the working surface were wiped with 70% Ethanol. The hood was left running in LAF purge mode for 10 minutes and then turned to red to normalise the flow for working.

**Step 1: Preparing 1 litre of Arabidopsis growth media:** 1Litre Duran bottle was labelled as M+S + 3% Sucrose, and with the date and the name of the operator and 4.4g of Murashige-Skoog media 30g sucrose (3% final concentration), 7g Agarose Type PGP, supplied by Park Scientific Ltd (0.7% final concentration) was weighed into a 1L Duran
1 litre of deionised polished water was added to the bottle. The growth media was mixed well in the solution by inverting the bottle three times. The solution was adjusted to pH 5.6 with 1M KOH. The lid of the Duran bottle was slightly loosened and a strip of autoclave tape was placed over the top of the lid. The bottle was autoclaved using the 2100 Classic Autoclave supplied by Jencons-PLS. (Autoclaving takes approximately 30 minutes but the programme was automatic and timing depended on whether the machine was starting from cold, or has recently been used). Once the autoclave programme finished, the Duran bottle was removed, the lid was tightened and placed in the Laminar Flow Hood to cool. (The following steps 2, 3 and 4 were carried out in a Laminar Flow Cabinet)

**Step 2: Pouring Agarose plates:** Using a black marker pen the outside of the bottom half of sterile petri dishes were labelled with details: M+S+3% Sucrose, name of the seed to be plated, date and name of the operator. Once the agarose was cooled enough to handle (usually 30-60 minutes after removal from the autoclave) it was poured directly from the bottle into the labelled triple vent 90mm petri dishes until agarose 2/3rds fill the bottom of the dishes. The petri dishes were pushed to the back of the flow hood and the lids of the petri dishes were left slightly open (offset by approximately 5mm), with the opened lid facing the filter at the back of the flow hood, to prevent condensation while minimising contamination. The plates were allowed to stand for one hour. Before carrying out sterilisation of seeds, 10% bleach solution was made in a 100ml Duran bottle using Parazone Jeyes Active Power Thick Bleach and deionised water (v/v). All the equipment needed viz., seeds, eppendorf tubes, bleach solution, sterile water, pipettes (P1000 and P20, pipette tips (1ml and 10-100µl Barky Ultipette Capillary tips), parafilm and scissors - were transferred to the Laminar Flow Hood.
Step 3: Sterilising Arabidopsis seeds: The seeds to be plated were transferred to an eppendorf tube labelled with the seed name. 1ml of 10% bleach (v/v in water) was added to sterilise the seeds. Each eppendorf tube was inverted five times in 3 minutes, to ensure all the seeds are exposed to the bleach. The seeds in bleach were left not for longer than 3 minutes as it would cause seed coat breakdown. The bleach solution was pipetted off and 1ml of sterile deionised water was added to remove residual bleach from the seeds. The tube was inverted five times to ensure thorough washing of the seeds. The water was removed with a pipette and repeated the sterile water wash. The majority of the water was removed, leaving a small volume in the base of the tube to facilitate plating of seeds.

Step 4: Plating out of Arabidopsis seeds: The lid of the petri dish was lifted, of the agarose plate to be plated, and angled towards the back of the flow hood, to help minimise contamination. Using capillary tips and a P20 Gilson Pipette, a 20µl aliquot of seeds was taken from the appropriate eppendorf tube. The pipette tip was inserted with the seeds under the lid to distribute the seeds individually at regular intervals over the plate. A maximum of 50 seeds were plate per plate. The plates were sealed by wrapping a 1cm strip of parafilm around the petri dishes where the top and bottom of the dishes meet. The petri dishes were place flat, in a cabinet where ensuring the conditions were set at 24 hours light and 22°C. The plates were left in the tissue culture cabinet 3 for 10 days or until the seedlings have 2-4 leaves.

Step 5: Recording Information: The seeds plated were listed on a record of Arabidopsis thaliana plant growth sheet. The sheet was completed with the following sections for referencing: description of the ecotype, seed name and number for identification and date plated.
Transferring Arabidopsis thaliana seedlings to soil

The seedlings were transferred 10 days after plating to the soil (75% medium grade peat, 12% screened sterilised loam, 3% medium grade vermiculite, 10% grit (5mm screened, lime free), 3.5kg osmocote plus ¾ month per m3, 0.5kg PG mix per m3, Lime to pH 5.5-6.0 and wetting agent). The plants are usually at the 2-4 leaf stage. On day of transferral the soil trays were made up according to number required. The soil is watered from above (from a height of approximately 50cm) using a hose with a rose attached, so as not to cause breakup of the soil surface. Once the surface looks wet some water was added to the bottom of the trays. Each tray was labelled with: the seed number, date of plating, and date of soil transfer, tray number and the name of the operator. The petri dishes containing seedlings were collected from cabinet where they were kept for germination. The seedlings were carefully picked off of the agarose plates with a sterilised cocktail stick and laid on the soil surface. Using the cocktail stick they were then gently pushed so that all the roots under the surface of the soil, leaving the seedling’s leaves resting on top of the soil. The trays were covered with an incubation lid, ensuring all vents are closed; the trays then were transferred to controlled environment growth room with growth conditions specific for Arabidopsis growth viz., Long day conditions: 16 hour day(Day- Light: 00:00-16:00 hrs, Humidity 75 %, Temperature 23°C; Night16:00-00:00 hrs, 80 %, 18°C).

The incubation lids were left on for 2-3 days, until seedlings have grown and look healthy, and removed when the plants were grown the required developmental stage for harvest. The date of transferral of seeds to soil was recorded on the relevant record of Arabidopsis thaliana growth sheet.
Harvesting of Arabidopsis thaliana plants for analysis

50ml centrifuge tubes were labelled with sample number, date of harvest and the operator's name. Separate tubes were used for different plant types and different trays of the same plant type. A minimum of three holes were made in the centrifuge tube caps with a pair of scissors. The plant samples were harvested at two growth stages namely rosette (5.10) and flowering (6.1-6.5) (Boyes et al., 2001). Harvesting of the plants were started at 14 hours into the growth day (2pm). A labelled 50ml centrifuge tube was dipped into a flask containing liquid. The plants were cut, using scissors, immediately below the rosette, as close to the soil surface as possible. The harvested plant was then put immediately using tweezers, into the centrifuge tube containing liquid nitrogen, so that it was frozen instantly, thus minimising any alteration of metabolites after harvest. Three to six plants were put in each tube. The cap of the tube was replaced, ensuring it has holes in it and immerse it in the flask of liquid nitrogen to ensure samples are kept frozen. The numbers of plants harvested were recorded for reference. The tubes of plant material was then transferred to a plastic bag labelled with date and operator name and stored at –80°C until preparation for analysis can be carried out.

Preparing Arabidopsis thaliana plants for NMR analysis

A pestle and mortar was chilled in a –80°C freezer for 30 minutes. A polystyrene box was filled with ice, a well was made in the ice and the mortar was placed in the well. Little Liquid Nitrogen was poured into the mortar and allowed to evaporate. The plant material from the –80°C freezer was removed and the plant material of the same ecotype and tray number were batched by emptying all centrifuge tubes containing the specified material into the pre chilled mortar. The liquid nitrogen was poured into the mortar and the pestle was used to grind the plant material. Grinding was continued until the material was fine and
homogenous. Any excess liquid nitrogen was allowed to evaporate off. A glass vial was taken, respectively labelled and the lid of vial was removed and the ground material was transferred to the vial, with a spatula. The material was retained at –80°C until further analysis.

**Analysing Arabidopsis thaliana plants by NMR**

**Step 1: Plant Extraction:** A solution of 0.05% TSP w/v 80:20 D$_2$O:CD$_3$OD (pH typically around 6.5) was made up. Three 1.5ml eppendorf tubes were labelled with the name of the plant material and the replications to be analysed. The material stored at -80°C was removed from storage and allowed stand on the laboratory bench for one hour, to allow the plant material to return to room temperature. 15mg ± 0.03mg of plant material were weighed out, into each labelled eppendorf tube. The water bath was turn set to 50°C and 90°C and allowed to reach the required temperature. The extraction method was a hot ethanol extraction and hence two temperatures were used for effective extraction of samples. To each eppendorf tube of plant material 1ml of the solvent, 0.05% TSP w/v 80:20 D$_2$O:CD$_3$OD, made in step 1.1 was added. Each tube was mixed, using a bench top whirl mix, until all the plant material was held within the solvent. The samples were then heated at 50°C in a water bath for 10 minutes. A second set of eppendorf tubes were labelled with the same descriptions as the first set in the water bath. The samples from the water bath were removed and centrifuged for 5 min in a bench top centrifuge. The supernatant were then transferred, using a 1ml Gilson pipette and a clean tip for each sample, to the clean, appropriately labelled eppendorf tube. The supernatant was heated at 90°C in a water bath for 2 minutes and the samples were removed from the water bath and transferred to a
refrigerator and store for 30 minutes. Clean NMR tubes were labelled with the same
descriptions as those on the eppendorf tubes. The samples were then removed from the
refrigerator and centrifuged for 5 minutes, in a bench top centrifuge. 750µl of the
supernatant was transfer, using a 1ml Gilson Pipette and a clean tip for each sample, from
each eppendorf tube to the appropriately labelled NMR tube and cap the tube. To get good
quality NMR spectra, there should be enough solution in the NMR tube to span the “active
volume” region. For a 5mm NMR tube the active volume was normally in excess of
500microlitres.750 µl were used because that was approximately the amount of clean
supernatant that can reproducibly collect from the extraction protocol.

**Step 2: NMR Analysis:** The samples were loaded on to the NMR instrument which was run
to the program set, selecting the appropriate solvent, (D$_2$O) and experiment (Watersup.lars 2K) and giving each sample a title. The data was collected by NMR acquisition Software
and then was used for further analysis. The parameters set for data collection is tabulated as
shown in Table 4.
Table 2.2 Parameters for NMR Data Collection

<table>
<thead>
<tr>
<th>Instrument Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>Bruker Biospin, Advance 400</td>
</tr>
<tr>
<td>Probe</td>
<td>5mm broad band multinuclear probe (BBO Z8248/0048)</td>
</tr>
<tr>
<td>Proton Frequency</td>
<td>400.1299232 MHz</td>
</tr>
<tr>
<td>Sample Introduction System</td>
<td>BACS</td>
</tr>
<tr>
<td>Automation Control Software</td>
<td>ICON-NMR</td>
</tr>
<tr>
<td>NMR acquisition Software</td>
<td>X win-NMR 9 (v3.5)</td>
</tr>
<tr>
<td>NMR tube size</td>
<td>5mm</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acquisition Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Set</td>
<td>Watersup.roth</td>
</tr>
<tr>
<td>Pulse Sequence</td>
<td>zgpr (low power pre-saturation pulse followed by high power 90° pulse)</td>
</tr>
<tr>
<td>Presaturation pulse</td>
<td>50.0 dB for 5 s</td>
</tr>
<tr>
<td>90° pulse</td>
<td>-6 dB for 7.9 µs</td>
</tr>
<tr>
<td>Sweep width</td>
<td>12.1084 ppm</td>
</tr>
<tr>
<td>Time Domain size</td>
<td>32k data points</td>
</tr>
<tr>
<td>Temperature</td>
<td>300 K</td>
</tr>
<tr>
<td>Irradiation Frequency/</td>
<td>4.7ppm</td>
</tr>
<tr>
<td>transmitter offset</td>
<td></td>
</tr>
<tr>
<td>Number of Scans</td>
<td>2048</td>
</tr>
<tr>
<td>Lock</td>
<td>Automated locking on D₂O</td>
</tr>
<tr>
<td>Sample Shimming</td>
<td>tune xyz (all under automation)</td>
</tr>
<tr>
<td>Sample Spinning</td>
<td>20 Hz</td>
</tr>
<tr>
<td>Receiver Gain</td>
<td>Optimized for each sample by the automation program au_zg (using command RGA)</td>
</tr>
<tr>
<td>other details</td>
<td></td>
</tr>
<tr>
<td>sample temperature</td>
<td>27°C</td>
</tr>
<tr>
<td>T1 values</td>
<td>not measured for the sample</td>
</tr>
<tr>
<td>Time Delay</td>
<td>5s (combines with 3.4s acquisition is enough for adequate relaxation)</td>
</tr>
<tr>
<td>spectral processing</td>
<td>0.5Hz line broadening</td>
</tr>
</tbody>
</table>
NMR Spectral processing with MestRe-C

A commercial processing software, MestRe-C, which is particularly designed to process NMR data is used (Figure 3). MestRe-C can display and manipulate single or multiple 1D spectrum and allows for detailed analysis and interpretation of NMR spectra. It is able to handle data formats from a wide variety of instrument vendors and it automatically recognises and converts data to the MestRe-C format when opened.

The data file taken from the spectrometer called the raw free induction decay (fid) were imported in the MestRe-C and all the fids for all the samples involved in the present study and their respective biological replications and technical replications were given a new appropriate name for identification (e.g. WTR1 refers to wild type, rosette, biological replication number 1.) Full Fourier Transformation (FT) was selected to get the proton spectrum onto the interface and to convert NMR signals from time domain to frequency domain which was done automatically with the help of option button on the software. Phase correction was done next on to the spectra on the screen to get the phase corrected by Automatic phase correction (Global method). Baseline correction was performed next on the phase corrected spectra. Two baseline corrections are used with a) multipoint baseline correction with linear picking and b) Auto baseline correction, to smooth the outliers to obtain a smooth baseline. Referencing was done to the spectra with Trimethylsilane (TMS) which was used as internal standard and was referenced to a tsp value of zero by expanding the spectrum and dragging the TMS icon manually and the spectra was calibrated (referenced). Bucket/binning, a commonly used technique for digitizing a spectrum into a row vector was done with a bucket width of 0.01 ppm and 0.04 ppm was applied, the spectra
was split into 1000 and 250 integrated buckets respectively and the resultant data (as absolute normalized values). The generated ASCII files were then imported to Microsoft EXCEL for the addition of labels saved for further analysis. All the samples were subjected to similar data processing procedures and all the resultant files are copied to a single Microsoft EXCEL spreadsheet and all the three technical replications for each single sample are averaged up to give rise to biological replications. The processed data was also saved as text files for statistical analysis.

Figure 2.1 Spectral processing with MestRe-C. (A) Raw NMR data exported into MestRe-C as FIDs;(B) Fourier transformation;(C,D) Phase correction before and after;(E) Auto-baseline correction (F,G) Baseline correction by linear correction before and after;(H) Calibration/Referencing with TMS shift value 0;(I) Bucketing Integration and integrals imported as ASCII files.

Data Analysis

The pre-processed from MestRe-C are then further altered by removing water regions(bins)450-560 from 0.01 ppm bucketed data and regions(bins)116-154 from 0.04 ppm bucketed spectra to eliminate any variability in suppression of the water signal.
The statistical data analyses and visualisation were performed using MeV. Two types of statistical analysis were done to the data; Analysis of Variance (ANOVA) and Hierarchical clustering (HCA). The statistical analysis ANOVA in MeV was carried out by uploading the data file and the numbers of groups was assigned followed by the P-value parameters and Bonferroni correction which gives a stringent condition. No correction gives a least stringent condition. The results are displayed in an open window which gives the values for the significant bins and non-significant bins. The significant bins after from ANOVA were chosen for performing clustering analysis of the data. HCA was chosen with Pearson's correlation with average linking clustering as linkage method. The results are clusters which were visualized as Gene Tree and heat maps. The schema of data processing and data analysis is given in Figure 2.2.

**Figure 2.2 Schema of the data analysis.** Wild type and four vtc mutants have two stages (rosette and flowering), two baseline correction (auto, linear), two band width (0.01 and 0.04 ppm), two normalizations with control (Total area and HPV (HPV=High Peak value) and statistically analyzed (HCA=Hierarchical clustering analysis, ANOVA=analysis of variation)
Chapter 3

RESULTS and DISCUSSION

NMR SPECTRAL FEATURES OF ASCORBATE-DEFICIENT MUTANTS

Four ascorbate-deficient mutants (vtc1, vtc2, vtc3 and vtc4) were employed in the current study. All mutants contained different levels of ascorbate as described in the Chapter 1 as these mutants have genetic defect in ascorbate biosynthesis pathways (Conklin, 1999; Conklin, 2000; Dowdle et al, 2007). These mutants are phenotypically different on the basis of their size, as previously found ascorbate is an important regulator of growth (Kercheve et al, 2011). The aim of the current study is to investigate the effect of ascorbate deficiency on metabolic changes analysed by NMR. NMR was used as an analytical tool to analyse the differences at metabolite level in two growth stages such as rosette and flowering in WT and vtc mutants. There are different techniques used for plant metabolic profiling as described in Chapter 1, but NMR spectroscopy is a very powerful technique which gives the structural elucidation of metabolites which are useful for identification. For the past many years, NMR has been found very useful technique because of its unique advantages viz., non-destructive, non-biased, easily quantifiable, requires no or little separation and needs no derivatization (Wisahart, 2008). NMR has played a central role in understanding the metabolic changes in different strains. The limitation of NMR is that it is unable to detect the low abundant metabolites and also provide no information about retention time (RT) and mass spectral.
Figure 3.1 represents the one dimensional H-NMR spectra from *Arabidopsis thaliana*. A total of 45 NMR spectra representing Wild type and four mutants of two different growth stages (rosette and flowering) were collected and spectral processed in a commercial software MestRe-C for exploratory analysis. The spectrum of both rosette and flowering shows a dominance of signals in the carbohydrate region and well defined signals both in aromatic and aliphatic regions (Fig 3.2a and 3.2b).

Figure 3.1 $^1$H-NMR spectra of the *Arabidopsis thaliana* (Wild type) extracted with in D$_2$O:CD$_3$OD. The bottom spectrum represents the original spectra and the spectra in the box shows the extended region of the original spectra.
Figure 3.2a ¹H-NMR spectra of *Arabidopsis thaliana* at rosette stage. Spectra here represent different *vtc* mutants. A: Wild type B: *vtc1*, C: *vtc2*, D: *vtc3*, E: *vtc4*.

Figure 3.2b ¹H-NMR spectra of *Arabidopsis thaliana* strains at flowering stage. Insert represents different *vtc* mutants. A: Wild type B: *vtc1*, C: *vtc2*, D: *vtc3*, E: *vtc4*. 
NMR DATA PRE-PROCESSING

Pre-processing of data is a major challenge for metabolomics study as spectral processing influences the outcome for the data (Schleif, 2007). The strains in the study were subjected to different steps normally involved in data pre-processing i.e., Fourier transformation, phase correction, baseline correction, peak alignment (referencing to internal standard) and bucketing (binning). Many of these steps are necessary, because the crude NMR data show artefacts due to physicochemical differences (Torgrip et al., 2006). (1) Two baseline correction in MestRe-C namely the auto baseline and linear baseline corrections were used in the present study. (2) Binning is another crucial step in data pre-processing, as it is an approach to solve the unmanageable dimensionality of the data and the inter individual differences in peak locations caused by slight variations in the sample environment (Meyer, 2008). Two different bin widths (0.04 ppm and 0.01ppm) were applied to divide the spectral regions into equal sized bins. The intensity values of each bin were integrated and annotated to the bin. The bins were then compared across the different spectra. (3) All the bin intensity were then normalised. Normalization is a row operation that is applied to the processed data from each sample and comprises methods to make the data from all samples directly comparable (Craig et al., 2006). (a) The binned data were normalized with total area where the bin values were converted as a fraction of the total spectrum area. (b) They were also normalized using highest peak value where the integrals were individually divided by the highest peak value (here the reference peak).

All the above mentioned parameters were applied in the present study while processing the NMR spectra and also after obtaining the processed data to see if there are any influences of these different parameters on the outcome of the data and to choose the best combination for further analysis.
MULTIVARIATE ANALYSIS OF ASCORBATE- DEFICIENT MUTANTS

Many approaches have been used for the analyses of NMR data. The most common methods used for the analysis of metabolites data are the multivariate data analysis such as principle component analysis (PCA) and the hierarchical clustering (HCA) clustering(Deferenez and Colquhoun, 2003). In the current study, the analysis of variance (ANOVA) was performed to analyse the variability of the data between different strains. Each strain contained three biological replicates and three technical replicates. The ANOVA was calculated on all the strains against WT (control) in flowering and rosette separately. Bonferroni correction was used, which kept the false positives to a minimum as hundreds of bins (integrated buckets) were being analysed. The p < 0.05 was selected as the significant values. The significant data (p < 0.05) from both stages of growth used in the current study was combined and subjected to HCA.

HCA clusters calculated the differences between biological replicates and strains. It also clustered features together which are behaving in a similar way on the basis of their chemical nature. HCA was performed using Pearson's correlation and average linkages, the most commonly used methods for the metabolomic and genomic analysis. The resulted dendrogram (rosette and flowering), which separates the data into branches and showed the variability between strains and the biological replicates. The horizontal (x-axis) represents the biological replicates of each strain and two growth stages while vertical (y-axis) shows the features (metabolites). The features which are similar in nature were clustered together as shown in Figure 3.3. The dendrogram showed all the biological replicates from rosette were clustered together but unfortunately, the biological replicates from flowering were far away from each other and not clustered properly (Figure 3.3).
This may be due to variability in biological replicates from flowering which needs further investigation. Interestingly, we observed a prominent difference in metabolites intensity between WT and the \textit{vtc} mutants. The differences were also observed in the flowering and the rosette in the accumulation of metabolites. The results indicated that spectra produced by NMR for metabolites intensity are able to primarily show difference between vegetative and flowering stages.

Sixty four combinations from two different growth stages were used, including two different baseline corrections, two bin widths, two different normalisations methods along with the non-normalised. The data were obtained after performing data pre-processing and post processing using different parameters as described earlier are presented in Tables 3.1-3.4.

As described before we observed variation between biological replicates in flowering compared to rosette. We are unable to investigate the variability between biological replicate in flowering in the current study which needs further work. So, for the further analysis here, we selected the rosette data which showed very close behaviour in biological replication.
Figure 3.3 HCA showing the difference in growth stages (rosette and flowering) and WT and the vtc mutants. Boxes and ellipses show the clustering of biological replicates in rosette stage the flowering stage respectively. M=mutants, R=rosette, F=flowering, WT= wild type. The scale shows intensity of colours from highest to lowest. Red = Highest intensity, Green = Lowest intensity. The significant difference was calculated by ANOVA (p < 0.05).
Table 3.1 Clustering analyses of all plant types in rosette stage with 0.01 bin width and different processing combinations. Cluster No = number of clusters where all the three biological replicates within and between strain clustered together, Strains = represents the strains in which are all the biological replicates are clustered close to each other.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>BIN WIDTH (PPM)</th>
<th>BASELINE CORRECTION</th>
<th>NORMALIZATION</th>
<th>CLUSTERING ANALYSIS</th>
<th>CLUSTERS (No)</th>
<th>STRAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>wt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>non-normalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>2</td>
<td>wt, vtc1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>3</td>
<td>wt, vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>wt</td>
</tr>
<tr>
<td>linear</td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>5</td>
<td>w, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>wt</td>
</tr>
<tr>
<td>HPV</td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td>ROSETTE</td>
<td>0.01</td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>non-normalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>3</td>
<td>wt, vtc 1, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td>auto</td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
</tbody>
</table>

| Total Area |                 |                      |               | HCA/ANOVA/Bonferroni | 2             | wt, vtc 1      |
|            |                 |                      |               | HCA/ANOVA/no correction | 2            | wt, vtc 1      |
|            |                 |                      |               | HCA/ANOVA/Bonferroni | 5             | wt, vtc 1, vtc 2, vtc 3, vtc 4 |
Table 3.2 Clustering analyses of all plant types in rosette stage with 0.04 bin width and different processing combinations. Cluster No = number of clusters where all the three biological replicates within and between strain clustered together. Strains = represents the strains in which are all the biological replicates are clustered close to each other.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>BIN WIDTH (PPM)</th>
<th>BASELINE CORRECTION</th>
<th>NORMALIZATION</th>
<th>CLUSTERING ANALYSIS</th>
<th>CLUSTERS (No)</th>
<th>STRAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>3</td>
<td>wt, vtc 1, vtc 4</td>
</tr>
<tr>
<td>ROSETTE 0.04</td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>2</td>
<td>wt, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>wt, vtc 1, vtc 2, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>3</td>
<td>wt, vtc 1, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>2</td>
<td>vtc 1, vtc 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/no correction</td>
<td>4</td>
<td>vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA/Bonferroni correction</td>
<td>5</td>
<td>wt, vtc 1, vtc 2, vtc 3, vtc 4</td>
</tr>
</tbody>
</table>
Table 3.3 Clustering analyses of all plant types in flowering stage with 0.01 bin width and different processing combinations. Cluster No = number of clusters where all the three biological replicates within and between strain clustered together, Strains = represents the strains in which are all the biological replicates are clustered close to each other.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>BIN WIDTH (PPM)</th>
<th>BASELINE CORRECTION</th>
<th>NORMALIZATION</th>
<th>CLUSTERING ANALYSIS</th>
<th>CLUSTERS (No)</th>
<th>STRAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>non-normalized</td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>1</td>
<td>wt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>1</td>
<td>wt</td>
</tr>
<tr>
<td>FLOWERING</td>
<td>0.01</td>
<td></td>
<td>non-normalized</td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>2</td>
<td>vtc 1, vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>1</td>
<td>vtc 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /no correction</td>
<td>1</td>
<td>vtc 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>2</td>
<td>wt, vtc 1</td>
</tr>
</tbody>
</table>
Table 3.4 Clustering analyses of all plant types in flowering stage with 0.04 bin width and different processing combinations. Cluster No = number of clusters where all the three biological replicates within and between strain clustered together, Strains = represents the strains in which are all the biological replicates are clustered close to each other. *Equal refers to the clusters which presented no hierarchy.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>BIN WIDTH (PPM)</th>
<th>BASELINE CORRECTION</th>
<th>NORMALIZATION</th>
<th>CLUSTERING ANALYSIS</th>
<th>CLUSTER S (No)</th>
<th>STRAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLOWERING</td>
<td>0.04</td>
<td>linear</td>
<td>non-normalized</td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>2 vtc 1, vtc 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HPV</td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>5 EQUAL *</td>
<td></td>
</tr>
<tr>
<td>Total Area</td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>2 vtc 2, vtc 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0 -</td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td></td>
<td></td>
<td>non-normalized</td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>2 vtc 1, vtc 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0 -</td>
<td></td>
</tr>
<tr>
<td>HPV</td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>2 vtc 1, vtc 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>5 EQUAL *</td>
<td></td>
</tr>
<tr>
<td>Total Area</td>
<td></td>
<td></td>
<td></td>
<td>HCA</td>
<td>1 vtc 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ ANOVA /no correction</td>
<td>2 vtc 1, vtc 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCA/ANOVA /Bonferroni</td>
<td>0 -</td>
<td></td>
</tr>
</tbody>
</table>
HCA was also performed on non-significant data (no ANOVA was performed here) as listed in the Tables 3.1 to 3.4. We observed the biological replicates were not clustered together, which makes sense because ANOVA is able to calculate the variability of the data and filtered out all the features which might be noise. Good clustering shows good experimental replication and consistency of the replicates (Ward et al., 2003). In general normalised data showed better clusters than non-normalized data and both the normalizations on the data were comparable in terms of clusters showing that the two different types of normalizations did not outperform each other, although total area normalization gave slightly increase in the number of clusters.

Definite differentiation was observed in the rosette stage where the separation between wild type and four mutants were obvious which also shows the reliability of all the three biological replicates and hence the rosette stage was chosen than flowering stage for further analysis to determine the influence of different baseline correction and bin widths on the data.

Four combinations from rosette stage were used to carry on further analysis: (1) Total area normalisation, (2) ANOVA (Bonferroni corrected bins), (3) auto and (4) linear baseline correction. The baseline correction was performed with 0.01 and 0.04 ppm bin widths. The results were listed in Tables 3.1 and 3.2. Auto baseline correction combination in 0.01 ppm bin width although comparable discriminated few more bins than linear baseline correction whereas they were equal in 0.04 ppm. It was concluded, that bin width 0.01 ppm discriminated more metabolites than 0.04 ppm, as bin width 0.01 ppm separated two to four times more integrated buckets (12-29 in numbers) than 0.04 ppm (7 in numbers).
Wild type and mutants classified themselves as five crisp clusters and in all the four combinations studied. In all four combinations chosen as mentioned before, the \textit{vtc}2 grouped itself as a separate group from the other strains (Figure 3.4). The separate grouping of \textit{vtc}2 suggested it shows the different behaviour in metabolic level, this may be due to lowest ascorbate level. As mentioned Chapter 1 the ascorbate levels in \textit{vtc} 2 mutants are 10-15\% of Wild type as compared to other which contained about 50-70\% for \textit{vtc} 3 and 4 (Conklin et al 2000; Dowdle \textit{et al}, 2007).

**Figure 3.4 Heatmap for the $^1$H NMR based metabolic profiles of \textit{Arabidopsis thaliana} Wild type and four ascorbate-deficient mutants in rosette.** The scale shows the colour intensity from highest to lowest. Red = Highest, Green = Lowest. Different colours represent different metabolites intensity which is significantly different between different strains. The wildtype cluster and its corresponding discriminating resonances (1.32, 1.33) are shown in ellipses.
DIFFERENTIALLY EXPRESSED BINS BETWEEN WT AND VTC MUTANTS.

The four combinations from rosette stage the bins separated by ANOVA and HCA were traced to their resonances and the list of all the resonances responsible for the separation of classes and groups at spectral level and are tabulated in Table 3.5. Major differences occurred in the regions 0-6 ppm which corresponds to the aliphatic and carbohydrate regions in the spectra. The regions where major resonances are discriminated correspond to the regions of amino acids and fatty acids (Ward et al. 2003). The unique resonances that showed visual differences across the strains from two bin widths and the HCA heatmaps were extracted and tabulated in Table 3.6. There are also some resonances which clearly might be from the same peak area. The bin width 0.01 ppm has also picked up resonances from carbohydrate regions than 0.04 ppm (e.g. 5.25 and 5.80 ppm).

The resonance 1.32, 1.33 that discriminated WT and other mutants were traced in the original spectra shows the presence of peaks (Figure 3.6). The intensity of both the resonances shows negative correlation to the ascorbate concentration in the vtc mutants and was proportional to the ascorbate deficiency with across strains (Figure 3.5). Identifying the spectral patterns forms the main focus rather identifying and quantifying metabolites. These resonances (features) can then used to identify the corresponding metabolites from the NMR database.
Table 3.5 List of resonances showing discrimination in the spectral regions.

<table>
<thead>
<tr>
<th>REGIONS (ppm)</th>
<th>RESONANCES (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00-2.00</td>
<td>0.96, 0.97, 0.98, 1.00, 1.04, 1.32, 1.33, 1.36, 1.48, 1.81, 1.88, 1.89 and 1.90</td>
</tr>
<tr>
<td>2.00-3.00</td>
<td>2.10, 2.11, 2.12, 2.13, 2.14, 2.15, 2.39, 2.40, 2.43, 2.45, 2.48, 2.46, 2.62, 2.71, 2.72, 2.91, 2.92, 2.93 and 2.94</td>
</tr>
</tbody>
</table>

Table 3.6 Resonances different between strains at spectral level with two bin widths. (Wt=Wild type, vtc=ascorbate-deficient mutants)

<table>
<thead>
<tr>
<th>STRAINS</th>
<th>0.01ppm</th>
<th>0.04ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>1.32, 1.33</td>
<td>1.36</td>
</tr>
<tr>
<td>vtc1</td>
<td>2.12, 2.13, 2.14, 2.15, 2.43, 2.45, 2.46, 2.47, 2.72, 5.80</td>
<td>0.99, 2.48, 2.72</td>
</tr>
<tr>
<td>vtc2</td>
<td>2.39, 2.72</td>
<td>2.40, 2.48, 2.72</td>
</tr>
<tr>
<td>vtc3</td>
<td>5.18, 5.80</td>
<td>2.40</td>
</tr>
<tr>
<td>vtc4</td>
<td>0.96, 0.97, 0.98, 1.00, 1.33, 2.39 and 2.94</td>
<td>0.99, 1.04</td>
</tr>
</tbody>
</table>

The resonances (1.32 and 1.33) were identified using the information from the previous literature. These resonances (1.32 and 1.33) putatively correspond to threonine and lactate respectively (Urenjak et al., 1993). Threonine is an amino acid in plants which is involved in defence mechanism with other defence related metabolites (Gonzales-Vigila, 2011). In plants, pyruvate, the end product of glycolysis, is converted to lactate by an enzyme called lactate dehydrogenase when oxygen is absent or in short supply (anoxia). Lactate provides the signal triggering for ethanol production and helps for prolonged survival in the absence of oxygen (Xia and Saglio, 1992, Tadege, et al, 1999). We are unable to investigate the nature of relationship between ascorbate and threonine and lactate at this stage. It needs further investigation.
Figure 3.5 The intensity of the resonances 1.32 and 1.33 ppm across the strains. Each dot represents the intensity of resonance. WT shows the lowest intensity than vtc mutants.

Figure 3.6 Identification of the resonances 1.32 and 1.33 ppm (shown as ellipses) in the original NMR spectra of one of the vtc mutant. The expanded regions are shown in the box. The insert is the NMR spectra of Wild type. The $^1$H-NMR-spectra shows very low signal in WT than vtc mutants.
Chapter 4

CONCLUSION

1 H-NMR spectroscopy has proved to be a valuable tool for unbiased metabolite profiling of Arabidopsis ascorbate-deficient mutants. ANOVA and HCA highlighted differences between the growth stages and strains. Combinations of different parameters were used to find the best method for the analysis of NMR-based metabolites data. The method was developed which allowed making multiple comparison such as two growth stages and five strains. The vtc2 mutants grouped separately from other strains which contained the lowest ascorbate suggested ascorbate deficiency might have some effect on the metabolic level. Two interesting resonances 1.32 and 1.33ppm were separated which were negatively correlated with the ascorbate levels in the vtc mutants. These resonances (1.32 and 1.33) are putatively identified as threonine and lactate which needs further validation by running standards.
Aharoni, et al. 2002

Abdel-Farid et al. 2009

Agnolet et al. 2010

Anterola and Lewis 2002

Bailey et al. 2003

Barros et al. 2010

Barth et al. 2004

Bino et al. 2004
Bochner et al. 2001

Blount et al. 2002

Brosche et al. 2005

Brown 2002

Cardoso-Taketa et al. 2008

Catchpole et al. 2005

Chen, et al. 2003

Cobas and Sardina 2003

Colville and Smirnoff 2008
Conklin, et al. 2006

Conklin and Last 1995

Conklin et al. 1999

Conklin et al. 2000

Cordoba-Pedregosa et al. 2003

Craig et al. 2006

Davey et al. 2000

Bennett 2005

Defernez and Colquhoun 2003

Defernez et al. 2004
De Gara et al. 1997

Delaglio et al. 1995

De Tullio 2004

De Vos et al. 2007

Dixon et al. 2006

Dowdle et al. 2007

Dunn and Ellis 2005

Eisenreich and Bacher 2007

Eskling and Akerlund 1997

Fiehn, et al. 2000a
Fiehn et al. 2000b

Fiehn 2002

Fiehn and Weckwerth 2002

Flamini et al. 2003

Fraser et al. 2000

Fryer 1992

Fukusaki and Kobayashi 2005

Gamache et al. 2004

Gander 1982

Garratt et al. 2005
Gatzek et al. 2002

Gautier et al. 2009

Gidman et al. 2003

Gonzales-Vigila 2011

Hadfield et al. 2001

Hall et al. 2008

Halket et al. 2005

Hamzehzarghani et al. 2005

Havaux 2003

Hirai et al. 2004
Hole et al. 2000

Huang et al. 2005

Ishikawa et al 2006a, b


Jackson and Mantsch 1996

Jackson et al. 2003

Jung et al. 2000

Katajamaa and Oresic 2007

Kercheve et al. 2011
Ketchum et al. 2003

Kikuchi et al. 2004

Kim et al. 2010

Kim et al. 2011

Kliebenstein 2004

Krishnan et al. 2005

Kristal et al. 2002

Kruger et al. 2008

Lange et al. 2001

Laing et al. 2007
Larkindale et al. 2005

Le Gall et al. 2001

Le Gall et al. 2003

Le Gall et al. 2005

Leiss et al. 2009

Li et al. 2008

Li et al. 2009

Linster et al. 2007, 2008

Linster 2008
Linster and Clarke 2008

Lisec et al. 2006

Liu et al. 2002

Loewus 1999

Lutterbach and Stockigt 1994

Lutterbach and Stockigt 1995

Luwe et al. 1993

Matsuda et al. 2010

Matsumoto and Kuhara 1996

Mochida et al. (2009)

Modarai et al. 2010
Merchant et al. 2006

Miller and Miller 2000

Noctor and Foyer 1998

Pallanca and Smirnoff 2000

Palsson 2000

Pastori 2003

Pavet et al. 2005

Pope et al. 1993

Prabhu et al. 1996

Ott et al. 2003
Rasmussen et al. 2006

Ratcliffe and Shachar-Hill 2001

Robinson et al. (2005)

Roessner et al. 2001

Roessner et al. 2000

Roos et al. 2004

Saito and Matsuda 2010

Saeed et al. 2006

Sato et al. 2004

Sauter et al. 1991
Schad et al. 2005

Schauer 2005

Schauer and Fernie 2006

Schleif 2007

Schöner and Krause 1990

Sharma and Davis 1994

Shulaev 2006

Smilde 2005

Smirnoff 2000

Smirnoff et al. 2001

Smirnoff and Wheeler 2000
Smirnoff 2011

Soga et al. 2002

Sumner et al. 2003

Suza et al. 2010

Tadege et al. 1999

Tagashira et al. 2005

Tikunov et al. 2005

Torgrip et al. 2006

Tunali et al. 2003
Tweeddale et al. 1998, 1999


Verdonk et al. 2003

Urenjak et al. 1993

van der Greef et al. 2003

Viant et al. 2003

Ward et al. 2002

Wang et al. 2009

Washko et al. 1992
Weichert et al. 2002

Wheeler et al. 1998

Wolfender and Hostettmann 1996

Wolucka et al. 2005

Wishart et al. 2008

Xia et al. 2009

Xia and Saglio 1992

Yamazaki 2003

Yang et al. 2007
Zhao et al. 2006