Investigating cell type specific metabolism using GFP as a reporter protein

Merja Rossi

Department of Plant Sciences

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2014

Supervisors: Professor R. George Ratcliffe and Dr. Nicholas J. Kruger

Abstract

Investigating cell type specific metabolism using GFP as a reporter protein

Merja Rossi, Wolfson College

Submitted for the degree of D.Phil, Trinity Term 2014

Metabolic flux analysis (MFA) is a powerful technique for quantifying the intracellular fluxes in central carbon metabolism. It relies on detection of stable isotope labelling from metabolites such as amino acids derived from protein. Current standard techniques are, however, unable to distinguish between different cell types in heterogeneous tissue. The aim of the thesis was to address this problem by developing and validating a strategy using green fluorescent protein (GFP) with cell type specific expression as a reporter protein for investigating the fluxes in specific cell types in the *Arabidopsis thaliana* root. The fundamental difficulty in applying a reporter protein strategy in a multicellular organism arises from the limited amount of recombinant protein expressed by the cells.

The main novel contributions of the work in this thesis are threefold. First, a robust protocol for purification of GFP from the roots of Arabidopsis seedlings and for detection of reliable mass isotopomer distributions from the amino acids derived from GFP are described. Secondly, the reporter protein strategy is validated in this biological system with a focus on showing the data obtained by the use of the reporter protein is equal to that normally obtained from the total protein fraction. To expand on this, stable isotope labelling in isolated root hair cells is explored. These cells are easily isolated and show potential as a model system for cell type specific metabolism. Finally, the experimental data provide evidence for the feasibility of measuring data from specific cell types with appropriate mass spectrometric techniques. Analysis of cell type specific gene expression in this system suggests differences in the primary metabolism of different cell types cannot be ruled out without further investigation. Based on small scale *in silico* modelling described in this thesis, new solutions capable of providing data on sub-populations of cells are required, if central metabolism of the cell types differs significantly.

Acknowledgements

I am indebted to my supervisors George Ratcliffe and Nick Kruger for providing me with the opportunity to work on this project and for their guidance, insight, patience and support throughout.

I would like to thank Monika Kalde for her help with setting up the immunopurification system and Shyam Masakapalli for all his help with metabolic flux analysis. I would also like to thank Markus Schwarzländer for providing seeds for the GFP expression line with constitutive expression and Sourav Datta and Liam Dolan for their help with the RSL4 over expression line and techniques related to root hairs.

I would like to express my gratitude to those who have donated their time, guided me with analytical techniques and allowed me to use their analytical instruments: James Wickens and James McCullagh from the Mass Spectrometry Research Unit at the Department of Chemistry, University of Oxford; Souvik Kusari and Marc Lamshöft INFU, TU Dortmund; and Pedro Bota and Lee Sweetlove at the Department of Plant Sciences, University of Oxford.

I would like to thank everyone at the Department of Plant Sciences, who have been a source of support, enthusiasm and expertise, but especially those who contributed through providing help, equipment, plant seeds or advice: Bernadette Gehl, Constantine Garagounis, Maurice Cheung, Gemma Hayton, Annette Kugler and Chaiyakorn Srisakvarakul. I acknowledge the financial assistance provided by the BBSRC.

Finally, I would like to thank my family and friends for their understanding, encouragement and invaluable support. I dedicate this thesis to my mother, Pirkko.

Abbreviations

ADP adenosine diphosphate

ATP adenosine triphosphate

BSA bovine serum albumin

CFP cyan fluorescent protein

DESI desorption electrospray ionisation

EDTA ethylenediaminetetra acetic acid

FACS fluorescence-activated cell sorting

FL fractional labelling

FTIRC Fourier transform ion cyclotron resonance

GC-MS gas chromatography mass spectrometry

GABA gamma-aminobutyric acid

GBP GFP binding protein

GFP Green fluorescent protein

GST glutathione S-transferase

LAESI laser ablation electrospray ionisation

LCM laser capture microdissection

LC-MS liquid chromatography mass spectrometry

LDI laser desorption/ionization

MALDI matrix assisted laser desorption/ionization

MFA metabolic flux analysis

MS Murashige and Skoog (medium) / mass spectrometry

MSTFA N-methyl-N-trimethylsilyl trifluroacetamide

NAD nicotinamide adenine dinucleotide

Ni-NTA Ni-nitrilotriacetic acid

NHS N-hydroxysuccinimidyl

NMR nuclear magnetic resonance

PCA principal component analysis

PI propidium iodide

RuBisCo ribulose-1,5-bisphosphate carboxylase/oxygenase

SD standard deviation

SIM selected ion monitoring

SIMS secondary ion mass spectrometry

tBDMS N-methyl-N-t-butyl dimethylsilyl

TCA tricarboxylic acid cycle

TFA trifluoroacetic acid

TOF time of flight

UPLC ultra performance liquid chromatrography

vMALDI-LTQ vacuum MALDI-linear ion trap mass spectrometry

YFP yellow fluorescent protein

Table of Contents

1 Introduction

1.1 The importance of developing strategies for investigating cell type specific metabolic activity

Central metabolism consists of cycles of core processes where the energy, carbon and reduction power of the cell are being regenerated and used in order to sustain life. The building blocks for the cells are derived from biosynthetic anabolic pathways and catabolic pathways, which break down nutrients and support necessary turnover of molecules. In the middle the reactions of central carbon metabolism link all of these main processes together. The machinery in this network consists of proteins: enzymes allow the flow of carbon through the system by taking part in catalytic reactions that build, combine and break carbon skeletons and remove and add phosphate or protons to cofactors.

This complex network can be studied in multiple ways: by quantifying enzyme kinetics, studying the transcript or protein levels, measuring the size of metabolite pools or looking at differences in gene expression. All of these investigations shape our understanding of the different layers of the metabolic network. However, they mostly work around the one aspect of the network that is maybe most characteristic of life: the metabolic activity of the cells (Sauer, 2006; Sweetlove et al., 2008).

Metabolic flux analysis is a powerful way of quantifying the intracellular fluxes in central carbon metabolism from experimental measurements. By following the flow of carbon through the system, it allows the investigation of flux phenotypes or "fluxomes" often represented by flux maps. In micro-organisms this approach has been highly successful leading to the ability to track absolute fluxes across the whole system with steady state stable isotope labelling experiments (Sauer, 2006). The success in micro-organisms has made MFA a desirable tool to transfer to work with more complex multicellular organisms. It could be argued it would be especially useful in plants, for the purposes of rationally engineering their metabolism towards production of biofuels, pharmaceutical compounds or beneficial nutrients in food crops (Sweetlove et al., 2008).

There are, however, limitations to application of steady state stable isotope MFA in plant systems. One of these limitations is the inability of standard methods to distinguish between different cell types in heterogeneous tissue. This inability arises from the use of metabolites that are common components in central carbon metabolism: since all cells contain these metabolites, it becomes impossible to distinguish between specific cell types from a mixture. The importance of this limitation depends on how different or similar the metabolism of different cell types is. If differences between cell types are small or they share a common pool of the metabolites of interest, it may be sufficient for most experiments to investigate aggregate fluxes obtained from average data. On the other hand, if the differences are significant, systems analysis approaches may need to be implemented at the level of different cell types. Developing strategies to measure the metabolic activity of sub-sets of cells therefore becomes important not only for finding biologically interesting differences, but for understanding cell type specific metabolism and its implications for MFA.

1.2 Central metabolism and importance of metabolic fluxes in plant systems

Earlier central metabolism was described as a network of enzyme reactions facilitating the flow of carbon through the network with an important role in cycling energy, reduction power and building blocks. Central carbon metabolism in plants produces precursors to build macromolecules that are used for different types of plant biomass such as cell wall components, RNA and protein. The secondary metabolism that it supports tends to branch out as more linear and less connected pathways (Kruger et al., 2012). In addition to providing the building blocks and precursor molecules for the cell, central carbon metabolism provides energy through a few main pathways: the oxidation of pyruvate in mitochondria through the citric acid cycle producing NADH for phosphorylation of ADP into ATP by the electron transport chain and fermentation in the cytoplasm under anaerobic conditions.

The citric acid cycle is not only an important node in terms of energy production, but is also central to production of amino acids in the cell. A few main metabolites involved in the TCA (tricarboxylic acid) cycle, namely pyruvate, alpha-ketoglutarate and oxaloacetate, are precursors for synthesis of glutamate, glutamine, arginine, proline, alanine, valine, leucine, isoleucine, aspartate, asparagine,

threonine and methionine (Figure 1). The carbon skeletons for the rest of the amino acids are derived from the pentose-phosphate pathway and phosphoenolpyruvate (phenylalanine, tyrosine and tryptophan), 3-phosphoglycerate (serine, cysteine and glycine) or ribose-5-phosphate (histidine). These amino acids act as precursors for further branches of metabolites and protein synthesis.

Metabolic fluxes represent the *in vivo* rates for metabolite interconversion in the central carbon network (Sauer, 2006). Changes in the flux distribution report on the metabolic phenotype of cells: increase or decrease in fluxes is linked to a shift in the carbon flow. Flux phenotypes can potentially also report on the functions of the system at a given time by elucidating which pathways are being used (Kruger et al., 2012).

The metabolic network is often described as a set of pathways that support specific functions of the cells. While the stoichiometry certainly provides the basis for cell metabolism, it seems this view might be somewhat limited. In his review of metabolic flux analysis, Sauer gives the pentose phosphate pathway as an example: while generally it is considered to provide precursors and redox equivalents to biosynthesis, its main function in many organisms seems to be to function as a catabolic pathway (Fuhrer et al., 2005; Sauer, 2006). Similarly Sweetlove and co-authors give an example of the tricarboxylic acid cycle in plants: in developing oilseed embryos the pathway is not cyclic, but shows a more linear flux to support export of citrate for biosynthesis of lipid (Schwender et al., 2006; Sweetlove et al., 2008). Fluxes, therefore, imprint the current metabolic phenotype of the cell on the blueprint of available pathways defined by the known stoichiometry of cell metabolism.

One of the main research questions related to fluxes in plant metabolism can therefore be defined as: what drives and regulates the fluxes? In micro-organisms shifts in fluxes can occur as a result of the cells trying to survive environmental stresses, which they are constantly exposed to. This responsiveness to external and intracellular signals is suggested to arise from different demands for energy, cofactors or precursors (Haverkorn van Rijsewijk et al., 2011). A set of regulatory mechanisms allows the cells to respond and adapt their central metabolism. These mechanisms can

include changes in the *in vivo* capacity of enzymes, indirect changes in substrates, products or allosteric effectors or regulation at the levels of transcription, translation and post-transcriptional and post-translational modifications (Chubukov et al., 2013). Each of these can have a different contribution to the overall flux phenotype. The regulation of a flux phenotype therefore consists of complex interactions between several regulatory layers, which can potentially only be understood through systems analysis approaches (Sauer, 2006; Sweetlove et al., 2008).

The engineering of plant metabolism for over production of desirable end products has mostly been unpredictable and in many cases unsuccessful (Allen et al., 2009a; Sweetlove et al., 2008). The most successful examples of metabolic engineering in crop plants have used several enzymes or transcription factors to control a range of biosynthetic reactions. Some examples of this are "golden rice" (Ye et al., 2000), improved anthocyanin levels in tomato (Butelli et al., 2008), nutritional enhancement through antioxidants (Enfissi et al., 2010), increased level of nicotine and benzylisoquinoline alkaloids (Sato et al., 2001), and improved folate levels (Diaz de la Garza et al., 2007). In addition, while it may be possible to increase flux through a pathway by over expression or introduction of a gene, it is also possible the increase in nutrients is counteracted by catabolic fluxes (for example engineering lysine content in barley seeds, (Brinch-Pedersen et al., 1996)). The ability to predict the flux phenotype in a system would allow more systematic approaches for metabolic engineering (Sweetlove et al., 2003).

The complexity of the regulation of flux in micro-organisms offers some perspective for why prediction of fluxes is difficult; the current level of understanding of the hierarchical control of plant metabolism limits the ability to predict flux phenotypes. Compared to micro-organisms, metabolic control analysis studying regulation of fluxes has not been carried out extensively and metabolic flux analysis approaches are more limited due to added complexity arising from compartmentation, considerations related to providing labelled substrates and heterogeneous tissues (Sweetlove et al., 2008). In multicellular organisms such as plants, there may also be complex connections between metabolism and other networks that deal with core functions such as development or stress responses (Allen et al., 2009a; Sweetlove et al., 2008).

Metabolic fluxes provide information on the metabolic activity of the cells at a level that is required for understanding the metabolic responses of plants to external and physiological demands (Kruger et al., 2012). This information is not available from metabolite pool size, since fluxes can be reversible or increase and decrease without effect on the pool size. Flux control arises from different layers of regulation, so there is no single static measure such as transcript or protein levels that can capture all the aspects of a flux phenotype without quantifying the fluxes in the system (Kruger et al., 2012; Sauer, 2006; Sweetlove et al., 2008).

1.3 ^{13}C metabolic flux analysis is a key diagnostic tool for studying fluxes in central carbon metabolism

Steady state stable isotope metabolic flux analysis (MFA) is currently the main technique for experimental measurement of fluxes. It relies on growing cells with a ^{13}C labelled substrate as the carbon source and detection of the isotope enrichment in metabolites at a time when the cells have reached a steady state of labelling. A mathematical analysis of the data measured with GC-MS or NMR against the known stoichiometry of the network allows deduction of fluxes in the system.

MFA is an established technique especially in micro-organisms, where prediction of the absolute fluxes in a whole organism has allowed use of the term "fluxome" (Sauer, 2006). Standard MFA methods measure fractional abundance of isotope labelling in biomass components such as cell wall, starch, sucrose and protein (Zamboni et al., 2009). Other molecules such as the intermediates of the TCA cycle can also provide labelling information that can be used for deducing fluxes (Rühl et al., 2012). However, they are often present at relatively low concentrations, have high turnover rates and can be analytically challenging due to their diverse chemistries (Sauer, 2006).

The most common method for ^{13}C MFA relies on detection of the stable isotope labelling from amino acids derived from protein, which is abundant in cells. Since amino acids are synthesised through different branches of central metabolism, they can provide labelling information across the metabolic network (Figure 1). In addition to this, labelling in other metabolites is usually analysed and further

biomass constraints are set by determining the consumption of substrate and biomass production (Masakapalli et al., 2010; Sriram et al., 2007; Williams et al., 2008).

The ^{13}C enrichment in metabolites can be analysed by NMR, which reports on positional isotopomers and by GC-MS, which provides mass isotopomer data (Ratcliffe and Shachar-Hill, 2006; Sauer, 2006). Normally approximately 10-15 amino acids can be detected from total protein hydrolysate with GC-MS analysis (Sauer, 2006). The amino acid peaks detected in spectra consist of different fragments arising from the ionisation of the molecules (Figure 2). Analysis of the mass isotopomer data derived from these fragments requires baseline correction (Lommen, 2009) and mass correction to correct for the presence of naturally occurring heavy isotopes in the derivatised fragment ions (Wahl et al., 2004). A combination of measurements using both analytical techniques increases the confidence in the flux estimates, but mass isotopomer data alone is sufficient to resolve fluxes in a metabolic network (Zamboni et al., 2009). Recently LC-MS/MS has also been used to detect mass isotopomer data (Mandy et al., 2014; Rühl et al., 2012), but the data may be affected by collision induced dissociation (Allen et al., 2014e). In the majority of recently published articles as well as this thesis, the focus was on using GC-MS, which offers higher sensitivity compared to NMR (Ratcliffe and Shachar-Hill, 2005).

The computational part of MFA results in deduction of fluxes from the mass isotopomer data and requires an accurate model of the metabolism of the target system. This consists of the reaction stoichiometry for each metabolic step that is included in the network. The experimental data in the form of mass isotopomer distributions and calculated errors is added to the model with any measured biomass constraints. The flux solution arises as a best fit solution from iterative fitting of the fluxes based on the experimental data and on sum squared standardised differences (the residuum) between the experimental and simulated data. If sufficient number of Monte Carlo simulations are run, the flux solutions with the lowest residua will represent the global minimum (best fit solution). The fitting process requires specialist software (13CFLUX or 13CFLUX2), capable of performing Monte Carlo simulations on the provided networks and data (Wiechert et al., 1999). The resulting best fit

solutions are collated and statistical analysis is applied on the fluxes. Finally, the results represent the estimated fluxes through the system in steady state and can be used in a flux map.

Another common way of reporting fluxes through a network is calculation of flux ratios (Szyperski, 1995). The advantage of flux ratios is that the method can be used with a more limited dataset (Fischer and Sauer, 2003; Rühl et al., 2011). Flux ratios are calculated using metabolite specific equations taking into account the labelling of a small number of metabolites. From this, the analysis infers relative contributions of the different converging pathways. In contrast, for metabolic flux analysis with 13CFLUX the minimum number of good quality independent isotopomer measurements is the number of unique free fluxes in the network. Under-determined networks lead to inability to resolve some of the fluxes, although this does not affect the fluxes that can be determined from such a network (Isermann and Wiechert, 2003; van Winden et al., 2001). When fluxes are indeterminable from the available data, it may be possible to define a set of measurements that can solve the problem. Alternatively, the network might have to be simplified (Kruger and Ratcliffe, 2009). The target of a metabolic flux experiment that aims to study the whole metabolic network is therefore to obtain a large set of data that allows estimation of the reliability of the fluxes.

1.3.1 Requirements for MFA

In addition to obtaining a large enough set of high quality data to allow resolving the fluxes in the network, there are three other main requirements for steady state stable isotope metabolic flux analysis. First, the ^{13}C label in the molecules must not affect the metabolic activity of the cells. At least one article has been published that found no detectable effect in intracellular fluxes from stable isotope labelling in plants (Kruger et al., 2007). Equally an example exists of work where kinetic isotope effects were found to introduce error at levels similar to error from the GC-MS measurements (Wasylenko and Stephanopoulos, 2013). It seems therefore that the potential error introduced is at a level that does not affect obtaining a global flux solution, but if taken into account might improve flux estimates. Secondly, the defined network must cover all the reactions involving carbon transition that affect the metabolites in the network. Finally, the system should reach a steady state of both

metabolism and isotopic labelling or the contributing unlabelled carbon sources should be taken into account in the model (Masakapalli et al., 2013; Wiechert and de Graaf, 1997).

1.3.2 Alternative approaches for metabolic flux analysis

Alternative metabolic flux analysis approaches include dynamic MFA, flux balance analysis and the developing field of non-stationary isotopic MFA. The first uses time course experiments to deduce fluxes in cultures which have not reached a pseudo steady-state by measuring rate of labelling introduced into metabolite pools (Heise et al., 2014), and the second is a computational approach, which uses the genome of an organism as a basis for a model predicting flow of metabolism through the metabolic network (Williams et al., 2010). The last technique has been developed in the last few years and relies on quantification of a metabolite's concentration at metabolic steady state and the transient ^{13}C labelling of free metabolites (Noh and Wiechert, 2011).

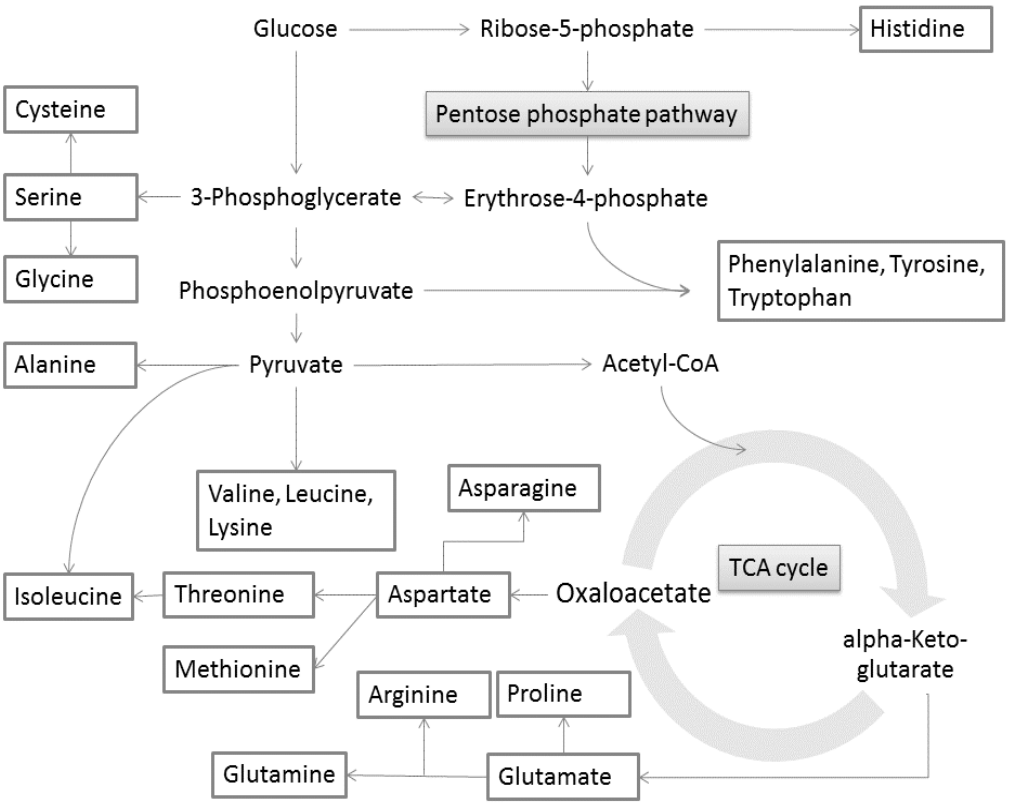


Figure 1: **Central carbon network with the precursors for amino acid biosynthesis shown.**

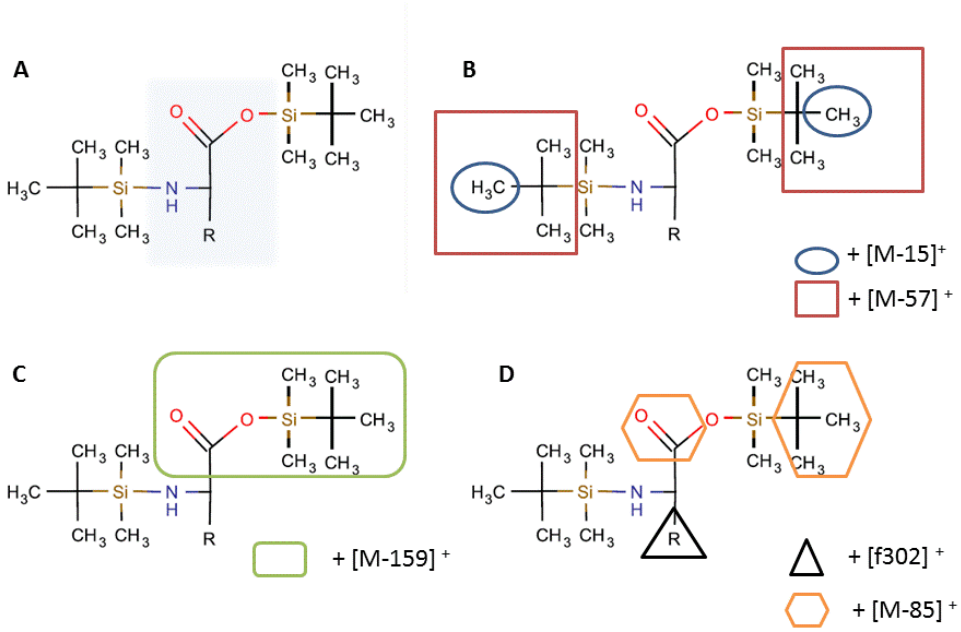


Figure 2: **Amino acid fragments resulting from ionisation of TBDMS-derivatised amino acids. A** The intact derivatised amino acid (highlighted) with its specific side chain R and two TBDMS-groups. The carbon skeleton of the amino acid is formed by the carbons in the highlighted area and the carbons from the TBDMS-groups are excluded. **B** Loss of methyl groups results in fragment $[M-15]^+$ and loss of the whole tert-butyl groups in $[M-57]^+$. The latter is often the most abundant fragment and therefore most commonly used for detection of amino acids. **C** Cleavage at the denoted position leads to a $[M-159]^+$-fragment and a C(O)O-TBDMS ion. **D** Loss of the amino acid specific chain R leads to the $[f302]^+$ double silylated fragment. Loss of the CO-group and a tert-butyl group results in the $[M-85]^+$-fragment. (Based on (Dauner and Sauer, 2000).)

1.4 Metabolic flux analysis in plant systems

Transferring steady state stable isotope MFA from micro-organisms to plants brings with it several new considerations. Firstly, plant cells are heavily compartmented and some parallel metabolic pathways are present in different compartments (Sweetlove et al., 2008). In order to produce flux maps of such systems, the pathways should be assigned to the correct compartment, which makes the computational analysis more demanding (Kruger et al., 2012; Masakapalli et al., 2014). Allen and co-workers have addressed the challenge of obtaining isotope labelling data from different metabolite pools by measuring labelling in amino acids from large and small sub-units of RuBisCo,

which are synthesised in the plastid and cytosol, respectively (Allen et al., 2012). Such reporters may not, however, be readily available in all instances. Subcellular fractionation has been suggested as another solution (Kruger et al., 2012). On one hand, MFA has already provided information on labelling in amino acid pools and contributions of parallel pathways in different compartments (Allen et al., 2012; Masakapalli et al., 2013), on the other, obtaining measurements from metabolites in different compartments presents a challenge.

Secondly, isotope labelling through a single carbon compound such as CO_2 is not an informative approach for traditional MFA, as it produces uniform labelling (Kruger et al., 2012). Systems where labelling is introduced solely through photosynthesis rely on measuring labelling time courses (Heise et al., 2014). In mixotrophic systems where photosynthesis may provide sugars into the system in addition to the labelled substrate, the introduction of unlabelled carbon can be quantified and taken into account or it may be enough to simply introduce an uptake flux in the model (Kruger et al., 2012). This has been done for example in cyanobacteria, although the flux analysis experiments were conducted in a pseudo steady state through continuous illumination (Alagesan et al., 2013). Most MFA experiments in plants use either ^{13}C labelled glucose (Masakapalli et al., 2010; Williams et al., 2008) or ^{13}C labelled sucrose (Lonien and Schwender, 2009) as a substrate and are limited to measuring a defined set of fluxes in central carbon metabolism. Choice of the most suitable substrate for MFA is, however, a consideration for each individual experiment and may help in obtaining a more meaningful flux map (Kruger and Ratcliffe, 2009).

Thirdly, steady state experiments require a metabolic and isotopic steady state. In tissues the metabolic steady state may be disturbed, for example through response to stresses. If these changes take place rapidly, the labelling in metabolite pools may not be at the corresponding isotopic steady state. Therefore most ^{13}C MFA experiments in plant systems have been conducted using cell cultures and excised tissues under continuous culturing conditions (Kruger et al., 2012). Use of such cultures also addresses the issue of providing the labelled substrate, which can be included as the carbon source in the growth medium.

Many studies have been conducted using cell cultures for example in Arabidopsis, (Masakapalli et al., 2010; Williams et al., 2008), tomato (Rontein et al., 2002), rice (Matsuda et al., 2007) and in tobacco callus tissue (Fernie et al., 2001). Other approaches include a range of heterotrophic plant tissues such as excised embryos (*Brassica napus (Junker et al., 2007; Schwender et al., 2003; Schwender et al., 2006)*, sunflower (Alonso et al., 2007a), soybean (Sriram et al., 2004), excised root tips (maize (Alonso et al., 2007b; Alonso et al., 2007c; Dieuaide-Noubhani et al., 1995)), cotyledons (soybean, (Iyer et al., 2008)) and hairy root cultures (*Catharanthus roseus* (Sriram et al., 2007)). In addition to studying different aspects of central carbon metabolism, effects of environmental signals on fluxes have been investigated: in Arabidopsis cell cultures oxygenation conditions and effects of nitrogen and inorganic phosphate have been studied (Masakapalli et al., 2013; Masakapalli et al., 2010; Williams et al., 2008).

While the main pathways and enzymatic reactions in the metabolic network in plants are well known, MFA has the capability to capture instances where the enzymes have deviated from their expected roles (Kruger and Ratcliffe, 2015). Some examples of MFA experiments that helped reveal such differences are pyruvate dissimilation requiring isocitrate lyase and carbon dioxide fixation in mycobacteria (Beste et al., 2011) and notably the discovery of the role of RuBisCo in refixation of CO_2 in fatty acid synthesis in green oilseeds, oilseed rape and soybean (Allen et al., 2009d; Schwender et al., 2004; Schwender et al., 2006). A limited number of studies have also been able to make use of MFA to elucidate substrate channelling (Graham et al., 2007; Young et al., 2011).

At the same time MFA experiments suggest central metabolism appears robust. For example, different oxygenation conditions in Arabidopsis cultures, starch deficiency and different substrates provided at different developmental stages in developing maize kernels or different growth conditions at different stages of the growth cycle in tomato suspension cultures did not cause significant changes in the flux distributions (Rontein et al., 2002; Spielbauer et al., 2006; Williams et al., 2008). The change in flux distribution depends on the effect on the total activity of the cell: if the ratio between input and output fluxes is affected, the flux distribution changes (Kruger and Ratcliffe, 2015). However, detectable changes have to be significant enough to be observed from the

measurement data and not be lost in the experimental error. Such changes have been observed as a result of changes in nutrient supply and biomass composition (Allen and Young, 2013; Iyer et al., 2008; Junker et al., 2007; Kruger and Ratcliffe, 2015; Masakapalli et al., 2014; Masakapalli et al., 2013).

MFA could provide a tool for metabolic engineering of plant metabolism, similar to that in micro-organisms. However, this role will also be affected by considerations arising from recent work: 1) modifications resulting in significant output may be too small to cause detectable change in the flux distribution, 2) the small changes that are being observed suggest engineering approaches might have to be subtle and 3) without systems level analysis the flux distributions measured from transgenic plants can be unpredictable and difficult to explain (Kruger and Ratcliffe, 2015). For this reason, it is likely that the role of MFA will be tightly linked to other modelling approaches, as well as systems level analysis of all the data available on different layers of the metabolic network. However, the unique ability to report on the intracellular activity of cells based on experimental data means MFA will make a significant contribution to the systems level analysis of metabolic networks.

1.5 Advances in investigation of the metabolism of specific cell types and single cells in plants

Investigation of metabolism at the level of cell types and single cells is limited mostly by three factors: the ability to isolate specific cells from the tissue without affecting their metabolic activity, limitations of techniques in adapting to small scale work and detection limits of techniques used. This is true for analyses of protein and gene expression, metabolomics and metabolic fluxes. Since the turnover times for metabolites can be extremely short, processing of samples after harvest should be carefully considered along with what effect the isolation or analysis has on the data acquired on metabolic activity (Fernie et al., 2011).

1.5.1 Isolation of cells from plant tissue

The main techniques used for isolating populations of cells from plant tissue include laser capture microdissection (LCM), fluorescence activated cell sorting of protoplasts (FACS), isolating of cell contents through a microcapillary and other types of mechanical disruption of tissue. LCM and

FACS are associated with processing times of up to 1.5 hours and require pre-processing of the tissue samples (Birnbaum et al., 2005; Espina et al., 2006; Nakazono et al., 2003). These two techniques have been most widely used to study specific cells and tissue types.

Applying cell isolation techniques to studying metabolic processes is not necessarily straightforward. As a result of applying standard LCM protocols, metabolites are extracted by reagents used as part of tissue fixation or washed out as part of embedding the tissue sample (Schad et al., 2005b). However, when tissues are coated with cryoprotectant to protect them from ice crystals forming between the plant cells, LCM can be done using cryosectioned tissue (Espina et al., 2006; Nakazono et al., 2003; Schad et al., 2005b). Another issue arises from the number of cells required to analyse metabolites, which cannot be amplified after collection. Collecting information from isolated cells for example on proteins is laborious: protein identification on gel requires isolation of 10,000-250,000 cells (Day et al., 2005). Schad and group found approximately 5,000 Arabidopsis vascular bundle cells were required to detect and quantify 68 metabolites with GC-TOF MS (Schad et al., 2005b). The same group also reported approximately 20,000 Arabidopsis vascular cells were required for identification of 33 proteins with a combination of LCM and LC-MS/MS and 250,000 cells for well resolved analysis by two-dimensional electrophoresis (Schad et al., 2005a). In addition to the high numbers of cells required for analysis of metabolites and proteins, difficulties in isolating small plant cells and issues with unspecific isolation due to the rigid cell wall may limit the feasibility of applying LCM methods for probing cell type specific metabolism in the Arabidopsis root (Nakazono et al., 2003).

In other plant tissues LCM has been successfully applied to studying cell type specific expression of genes and building cDNA libraries: a study in *Heliotropium indicum* confirmed cell type specific expression of genes encoding for pyrrolizidine alkaloids in lower leaf epidermis and epidermis of the stem (Sievert et al., 2015) and genes expressed differentially in epidermal cells or vascular cells as well as in the compartments of the kernel have been identified in maize (Nakazono et al., 2003; Zhan et al., 2015). Another interesting example of application of LCM is the characterisation of transcriptomes of outer and inner epidermal layers, collenchyma, parenchyma and vascular tissues

of the tomato fruit pericarp (Matas et al., 2011). This investigation provided insight into spatial distribution of regulatory and structural genes related to energy metabolism, secondary metabolism and cell wall biology. The tissue type specific patterns of expression also helped identify a cuticle on the inner epidermis of the pericarp.

A large number of studies on sub-populations of cells from the Arabidopsis root have used fluorescence activated cell sorting to isolate cells from Arabidopsis lines with cell type specific GFP expression. FACS of protoplasts derived from the Arabidopsis root with cell type specific GFP expression relies on enzymatic digestion of the cell wall followed by cell sorting (Birnbaum et al., 2005). Some examples of such work include a gene expression map (Birnbaum et al., 2003), protein expression map (Petricka et al., 2012a) and a metabolite map (Moussaieff et al., 2013) of the cell types in the Arabidopsis root. The advantage of FACS is the ability to provide large numbers of cells of a specific type to be combined with a variety of analytical techniques. However, the effects of protoplasting on the data obtained from these experiments must be taken into account: Birnbaum and colleagues found that 356 genes were induced by protoplasting in a microarray dataset used for analysing gene expression in Arabidopsis roots (Birnbaum et al., 2003).

Additionally information on specific cell types has been obtained by isolating contents of cells with microcapillaries. These experiments have been carried out both with and without GFP markers to locate the cells of interest (Brandt et al., 1999; Karrer et al., 1995). Other ways to mechanically isolate cells include dissecting microsamples for spatial analysis (Kajiyama et al., 2015) and removing a layer of cells by tape followed by disruption of the tissue and sorting out vasculature and epidermis cells (Endo et al., 2014). The latter technique yielded novel information on tissue specific aspects of the circadian clock in the Arabidopsis leaf: asymmetrically coupled clocks appear to be regulating each other in vasculature and mesophyll cells.

1.5.2 Mass spectrometric approaches for analysis of single cells and cell types

In recent years mass spectrometric detection of metabolites and proteins directly from tissue surface has become possible due to development of new types of techniques for biological sciences. One of

these techniques is mass spectrometric imaging (MS-imaging). MS-imaging techniques aim to detect spatial distribution of molecules in a tissue sample by desorption, ionisation and subsequent MS-analysis. Like microdissection, several of these techniques require the sample to be fixed or otherwise processed. Washing steps and application of matrix have an effect on what molecules can be detected from the sample surface (Burrell et al., 2007).

One of the most widely used techniques is matrix assisted laser desorption ionisation (MALDI) mass spectrometry. MALDI-imaging has been widely used in other biological systems for studying spatial distribution of proteins and with on tissue tryptic digest can be combined with peptide analysis. In recent years this technique has been adapted for use in plant tissues. In addition to application of a suitable matrix that allows detection of the compounds of interest, MALDI-imaging of plant tissue normally requires cryosectioning (Grassl et al., 2011). The imaging of small metabolites has so far been limited by a high background in the low mass regions and difficulties with matrix crystal formation (Takahashi et al., 2015). In addition, root samples can require specific protocols for sample preparation due to their water content (Kaspar et al., 2011).

Nevertheless, MALDI-imaging of a variety of metabolites is increasingly possible from different types of plant tissues. Li and group alone used atmospheric pressure infrared MALDI imaging to detect over 50 small metabolites from a variety of plant organs from several different plants (Li et al., 2008b). In wheat seeds MALDI-imaging of mass charge values corresponding to ions derived from arginine and sucrose or glucose-6-phosphate and sucrose revealed very different spatial distribution of the two metabolites corresponding to different parts of the embryo (Burrell et al., 2007). Structural and spatial variability of polysaccharides in the cell walls of wheat grains have also been studied (Velickovic et al., 2014). MS-imaging of cotton seeds revealed distinct heterogeneous distribution of triaglycerols and phosphatidylcholines with several other lipids also detected at cellular resolution (Horn et al., 2012). The wax esters and hydrocarbons on Arabidopsis and date palm tree leaf samples on the other hand showed a homogenous distribution (Vrkoslav et al., 2010). Cha and colleagues describe detection of a range of secondary metabolites with colloidal graphite-assisted LDI MS imaging from Arabidopsis tissues. Their experiment revealed tissue-specific

accumulation of flavonoids in flowers and petals (Cha et al., 2008). GABA has also been found to accumulate in a specific location in eggplant seeds (Goto-Inoue et al., 2010). Despite the wide spread use of MALDI-imaging in analysis of proteins and peptides in other biological systems, protein and peptide analysis with this technique in plants has so far been limited to only a few instances (Cavatorta et al., 2009; Kaspar et al., 2011).

The success of MALDI-imaging has led to many slightly modified techniques for imaging of plant tissue. An exciting recent development is the LDI-imaging experiment carried out in Arabidopsis root tip by Takahashi and colleagues: roots or etiolated seedlings were glued on to indium tin oxide-coated conductive glass, vacuum dried and coated with suitable matrix or analysed without matrix with an LDI-Fourier transform ion cyclotron resonance (FTIR)-MS. The modifications applied by the group allowed imaging of putative metabolites from 4-day old seedlings. Another technique with improved spatial resolution is vacuum MALDI-linear ion trap mass spectrometry (vMALDI-LTQ). This technique has allowed MS-imaging of single pollen grains on Arabidopsis flowers and identification of lipid metabolites on surfaces of Arabidopsis roots, including root hairs (Jun et al., 2010).

Not all MS-imaging techniques require application of a matrix. One of these techniques is desorption electrospray ionisation (DESI) mass spectrometry. This technique is based on impact of a stream of charged solvent droplets on the tissue surface. This type of extraction of metabolites from the surface of the tissue is, however, particularly problematic for analysis of plant leaf tissue due to the wax on the surface affecting this interaction (Janfelt, 2015). Instead, the tissue is imprinted on a surface, which can be analysed. This type of DESI-imaging has been used for example to produce images of chlorophyll degradation (Müller et al., 2011). Laser ablation electrospray ionisation (LAESI) mass spectrometry is another example of direct mass spectrometric analysis from the surface of tissue that has been used for metabolic profiling of single cells from *Allium cepa* (Shrestha and Vertes, 2009) and for cell-by-cell imaging of the pigment ion cyanidin (Shrestha et al., 2011). A variety of oligosaccharides were detected and in addition pigmented cells were found to contain anthocyanins and other flavonoids, which differentiated them from adjacent cells. Finally, high-resolution

secondary ion mass spectrometry (SIMS) is a mass spectrometry technique with the capability to report on selected ions in individual cells and even subcellular distribution of trace elements. It has been used for example to study resource flow in soil-plant systems through quantitative spatial imaging (Kilburn et al., 2010). While nanoSIMS has limited capability to report on larger molecules, it can image the spatial distribution of stable isotopes such as ^{13}C with high resolution (Kilburn et al., 2010). NanoSIMS has already been used to image $NaH^{13}CO_3$ in the coral *Galaxea fascicularis*, which demonstrates the potential for use in applications requiring detection of labelled compounds (Clode et al., 2009).

Resolution and high background produced through MS-imaging currently often limit the application of these techniques for MFA. However, they show promise for studying spatial distribution of metabolites or for example pesticides (Mullen et al., 2005) and pathogen infection (Hamm et al., 2010; Tata et al., 2015). In the future the development of these techniques may lead to the sensitivity required for detecting isotope labelling either from peptides, soluble amino acids or intact metabolites.

1.6 Reporter protein strategy for studying cell type specific metabolism

Reporter proteins have great potential in MFA work as reporters on fluxes in central carbon metabolism: they consist of component amino acids, which are synthesized through different branches of the metabolic network. By breaking the reporter protein into smaller peptide fragments or amino acids, it is be possible to measure the isotope enrichment in these components. If two proteins or protein fractions are synthesised using the same pool of amino acids, purifying one of them is sufficient to report on the labelling in the total protein fraction. The main problem with such methods so far has been to find a highly specific purification method for the reporter protein so it can be efficiently extracted from the total protein fraction and subsequently obtaining sufficient amounts of the protein for consistent detection of the isotopic enrichment in the amino acids or peptides through mass spectrometry.

Two examples of work with reporter proteins in plants have already been described in this chapter: the use of RuBisCO sub-units to report on metabolic information in different compartments (Allen et al., 2012) and the technique described by Mandy and colleagues (Mandy et al., 2014) that aimed to report on labelling in peptide fragments. The examples that demonstrate the potential of the reporter protein most efficiently are from work with micro-organisms.

The first example of work where a foreign protein was expressed in bacteria in order to report on the labelling in the total protein fraction used transgenic bacteria expressing GFP (Shaikh et al., 2008). Shaikh and colleagues showed that the labelling patterns detected from the total protein fraction and the His-tagged GFP protein were equivalent (Shaikh et al., 2008). This experiment suggests the two protein fractions, total protein in the cell and the foreign GFP protein, were synthesized using the same pool of component amino acids, which is an important requirement for the use of a reporter protein strategy. The same group later used the strategy to study protein expression in lag phase (Shaikh et al., 2010).

Rühl and colleagues repeated the experiments on validation of GFP as a reporter protein, but in addition they purified GFP from a mixture of wild type bacteria and a sub-culture with GFP expression (Rühl et al., 2011). Although they reported limitations on the method especially in terms of amount of protein that needed to be purified from the sub-culture, they were able to deduce flux ratios using the data. Further evidence of the ability of GFP to report on intracellular fluxes from mixtures was provided by detection of altered fluxes in a sub-culture using data derived from the reporter protein. Transgenic bacteria expressing GFP and with deletions in the TCA cycle were grown in sub-cultures with wildtype bacteria and flux ratios were derived from the ^{13}C data derived from the GFP. This work again highlighted some of the challenges in cell type specific analysis, but it also successfully compared flux ratios from the two deletion mutants to the wild type based on the data obtained from the purified GFP. Very recently work has also been published on using photosystem I as a reporter protein for ^{13}C labelling when co-culturing cyanobacterium with heterotrophic bacteria (You et al., 2015).

The reporter protein strategy may be very useful for studies in bacteria for example in investigations of biofilms or in studying host pathogen interactions, but it also renders itself for use in multicellular organisms, which present the same problem of isolation of specific components from a mixture. In this work the reporter protein strategy is validated for use in the roots of *Arabidopsis thaliana*. The use of lines expressing GFP at only one of the cell types allows purification of protein that is only present in a single cell type at a time from a mixture, therefore bypassing the issues with effects of isolating single cells from the organism.

1.7 Choice of experimental system

The Arabidopsis seedling cultures used in this work satisfied the main requirements for metabolic flux analysis: the feeding of labelled substrate could be done by providing it in the growth medium and the culture conditions could be kept continuous for the system to reach and maintain isotopic and metabolic steady state. In addition to the GFP expression lines for several of the cell types in the root being readily available, there is also published data available on the gene expression of the key cell types that were chosen for this study: stele, cortex, endodermis and epidermis. The root hair cells represented an opportunity to further validate the technique and especially to investigate the differences between wild type root hairs and the root hairs growing under over expression of the RSL4 transcription factor, known to cause the longer root hair phenotype.

1.8 Statement of aims

The aims of the work presented in this thesis were to develop the current techniques used for steady state stable isotope metabolic flux analysis (MFA) by validating a reporter protein strategy for investigating cell type specific metabolism in the *Arabidopsis thaliana* roots and to explore the feasibility of applying it to report on cell type specific metabolism in different cell types in this system. Standard methods for MFA derive the required labelling information on amino acids from the total protein fraction, but individual proteins can equally report on the ^{13}C stable isotope labelling (Allen et al., 2012; Mandy et al., 2014; Rühl et al., 2011; Shaikh et al., 2008), as long as they are synthesised form the same pool of precursor amino acids (Allen et al., 2012).

The main objectives were to first address known issues around detection of ^{13}C enrichment from small amounts of amino acid fragments from a biological matrix with GC-MS and secondly to show that the reporter protein reliably reports on the same stable isotope labelling as the total protein fraction. Using GFP purified from specific cell types, the third objective was to address the question: Can the mass isotopomer distributions in amino acids derived from the reporter protein with cell type specific expression in the Arabidopsis root be reliably measured and if so, what can we learn from investigating cell type specific metabolism through the reporter protein strategy?

Metabolic flux analysis has contributed to the current understanding of metabolic activity of cells through reporting on the intracellular activity of cells. In Chapter 3 simple network models provide evidence that flux estimates from labelling data obtained from a mixture of cells with different flux distributions might not faithfully report on the average fluxes in the system. If flux distributions differ significantly between cell types in a multicellular system, studying metabolic fluxes in the system requires new strategies capable of reporting on stable isotope labelling in metabolites in these sub-populations of cells.

One such solution is provided by the reporter protein strategy. The validation of the strategy is comprised of two parts: first in Chapter 4 highly specific purification of GFP and development of protocols for detection of isotope labelling from constitutively expressed reporter protein and secondly, validation of the use of the reporter protein in the biological system in Chapter 5. The latter consists of confirming 1) that the isotopomer distributions detected from the reporter protein are equal to those detected from the total protein fraction, 2) that GFP expression is not affecting the data and 3) what effect photosynthesis has on the labelling detected from hydroponically grown Arabidopsis seedlings. Together these steps form the basis for use of the reporter protein strategy in this system.

Finally, Chapter 6 and Chapter 7 focus on work with specific cell types. In Chapter 6 isolated root hair cells are investigated as a potential model system for work on cell type specific metabolism: they are easily isolated, which makes them an ideal cell type for studying metabolic activity both

with standard methods and with strategies that provide more limited datasets. Chapter 7 examines the feasibility of measuring mass isotopomer distributions from amino acids derived from the low amounts of GFP that can be purified from Arabidopsis lines with cell type specific expression of the reporter protein in stele, cortex, endodermis and epidermis cells. Analysis of cell type specific gene expression data on genes linked to central metabolism demonstrates that significant differences in the metabolism of these cell types cannot be ruled out without investigating the metabolic activity of the cells further.

2 Materials and methods

2.1 Chemicals, spin columns, resins and kits

For stable isotope labelling experiments [$^{13}C_6$]-glucose was purchased from Goss Scientific Instruments (Crewe, United Kingdom). The derivatisation reagent TBDMS (N-methyl-N-(t-butyldimethylsilyl) trifluoroacetamide + 1% t-butyl-demithylchlorosilane) was purchased from Regis Technologies (IL, USA) or Sigma-Aldrich (MO, USA) and MSTFA (N-methyl-N-(trimethylsilyl) trifluoroacetamide) from HiChrom (Reading, United Kingdom) or Sigma-Aldrich. Complete protease inhibitor tablets were purchased from Roche (Basel, Swtizerland). General chemicals were purchased from Sigma-Aldrich or Merck Millipore (MA, USA).

Ni-NTA (nickel-nitrilotriacetid acid) agarose and kits for plasmid purifications and gel extractions were purchased from Qiagen (Hilden, Germany). N-hydroxysuccinimide-activated sepharose was purchased from GE Life Sciences (Chalfont St Giles, United Kingdom). Amicon Ultra 0.5 ml with 10K ultrafiltration membrane and solid phase extraction C_{18} ziptips were purchased from Merck Millipore and Vivaspin protein concentrators from Thermo Fisher Scientific (MA, USA).

Restriction enzymes and DNA molecular size ladder were purchased from New England Biolabs (MA, USA). Prestained protein molecular weight ladder was purchased from Thermo Fisher Scientific. Commercial GFP nanotrap agarose beads and a 96-well microplate coated with GFP binding protein were purchased from Chromotek (Planegg-Martinsried, Germany). Coomassie Brilliant Blue was purchased from Bio-Rad (CA, USA).

2.2 Software

Specialist software used for data processing included: 13CFLUX for simulations of isotopomer data and Monte Carlo simulations of flux solutions (Wiechert, 2001), available free for academic use from http://www.13cflux.net), MetAlign software for baseline correction (Lommen, 2009), available for academic use from https://www.wageningenur.nl/en/show/MetAlign-1.htm), AMDIS for identification of metabolites derivatised with TMS from the NIST metabolite database, MATLAB with the mass correction tool MSCorr (Wahl et al., 2004), Agilent Chemstation (for

spectra obtained from Agilent 5975C in .D format), Agilent Mass Hunter (for spectra obtained from Agilent 7200 in .D format), Mass++ (Tanaka et al., 2014b) and Xcalibur (for spectra obtained from Thermo Fisher IQS in .raw format).

2.3 Analytical equipment

GC-MS instruments used were Agilent 7890A GC coupled to 5975C single quadrupole mass detector, Trace GC Ultra from Thermo Fisher coupled to an ISQ mass spectrometer and Agilent 7890B GC coupled to 7200 Q-TOF mass detector. A Beckman Coulter DTX880 microplate reader was used for fluorescence measurements and a Zeiss LSM510 confocal microscope for confocal images.

2.4 Plant material

2.4.1 *Arabidopsis thaliana* lines

Arabidopsis thaliana lines were kindly provided by colleagues at the Department of Plant Sciences, University of Oxford: the wildtype used was *Arabidopsis thaliana* L. (Heynh) 'Columbia' ecotype, the line with constitutive GFP expression was roGFP expression line with GFP expression in the cytosol (Schwarzlander et al., 2009), the lines with cell type specific expression of GFP were pWEREWOLF::GFP Lee and Schiefelbein, 1999; Sena et al., 2004), pWOODEN LEG::GFP (Mahonen et al., 2000), pSCARECROW::GFP (Birnbaum et al., 2003), pCORTEX::GFP (Lee et al., 2006), pET111::GFP (Nawy et al., 2005) and the line with root hair specific GFP expression and RSL4 over expression (Yi et al., 2010).

2.4.2 Hydroponic cultures of *Arabidopsis thaliana* seedlings

Arabidopsis thaliana ecotype Columbia and transgenic lines with GFP expression were grown in hydroponic cultures with 25 ml of liquid medium in 100 ml wide necked conical flasks with the top covered with aluminium foil (Figure 3). 30-50 seeds were sterilised in 1.5 ml microcentrifuge tubes by adding 1 ml of 70% ethanol for two minutes after which ethanol was discarded and 1 ml 50% bleach (12-14% sodium hypochlorite diluted 1:1 with water) was added. The tubes were inverted every two minutes for ten minutes before the bleach was discarded. The seeds were washed five

times with 1 ml sterile deionised water and as much water as possible was discarded after the final wash. The seeds were transferred into conical flasks containing the growth medium that had been autoclaved in flasks (120 °C for 10 minutes 15 psi) by washing the seeds out of the microcentrifuge tubes with the medium under sterile conditions. The flasks containing medium and seeds were then kept at 4°C on an orbital shaker (at 70-100 rpm) for two days before being moved to the growth rooms.

Seedlings were grown for 14 days on an orbital shaker at approximately 70 rpm under a 16 h light and 8h dark cycle at 22°C. The root and leaf tissue was then separated and samples were either frozen in liquid nitrogen and stored in a -80°C freezer or used on the day of harvest.

Figure 3: **Hydroponic Arabidopsis thaliana culture after 14 days of growth.**

MS medium for experiments on lines expressing GFP

Half strength Murashige and Skoog (MS) medium supplemented with 0.05% (w/v) 2-[N-morpholino]ethanesulfonic acid (MES)-KOH buffer and 1% (w/v) glucose was used as the standard growth medium with the pH of the medium adjusted to 5.7-5.8 before autoclaving. For experiments that required isotope labelling, 20% of the glucose in the medium was provided as [$^{13}C_6$]-glucose and 80% as unlabelled glucose (at isotopic natural abundance).

Johnson's medium for experiments on root hairs

To induce the longer root hair phenotype in the RSL4 over expression line, the medium described by (Johnson et al., 1957) as modified by Ma and colleagues (Ma et al., 2003) and used in the experiments on RSL4 over expression line (Yi et al., 2010) was used instead of MS medium for all seedlings that were used in the experiments on root hairs. The medium consists of 3 mol m^{-3} KNO_3, 2 mol m^{-3} $Ca(NO_3)_2$, 0.5 mol m^{-3} $MgSO_4$, 25 mmol m^{-3} KCl, 12.5 mol m^{-3} H_3BO_3, 1 mol m^{-3} $MnSO_4$, 1 mmol m^{-3} $ZnSO_4$, 0.25 mmol m^{-3} $CuSO_4$, 0.25 mmol m^{-3} $(NH_4)_6Mo_7O_{24}$, 25 mmol m^{-3} Fe-EDTA, 0.55 mol m^{-3} myo-inositol and 2.5 mol m^{-3} MES. The sucrose in the medium (29.2 kmol m^{-3}) was replaced with glucose to allow stable isotope labelling and Phytagel was left out to allow use in liquid cultures. Since all of the phosphorus in the medium is normally provided in the Phytagel, 1 mmol m^{-3} $NH_4H_2PO_4$ used in the published work to produce different phosphorus concentrations was added to the medium. These two changes were the only difference to the medium used in studies in which the longer root hair phenotype was originally described (Yi et al., 2010). The pH of the medium was adjusted to 5.7.

2.4.3 Experiments on root hair exudates

Seedlings were grown in Johnson's medium for 14 days as described above. The seedlings were then rinsed with deionised water and transferred to 100 ml flasks containing 25 ml of sterile deionised water only. The growth conditions were otherwise identical and the seedlings were kept in the water cultures for a further three days. After this the plant material was discarded and the water was collected and freeze dried. The dry material was resuspended in 1 ml water. A 200 µl fraction of the

total volume was dried in a speed-vac with internal standard (ribitol 60 μg) added, derivatised with MSTFA and analysed with GC-MS.

2.5 Extraction and hydrolysis of protein from seedling cultures

2.5.1 Extraction of total protein fraction

Plant tissue harvested from seedling cultures at 14 days of growth was patted dry with tissue and ground into powder in liquid nitrogen. The powder was extracted with 5 ml of 100% methanol and the remaining insoluble material was homogenised in 2 ml of extraction buffer containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-KOH (pH 7.5), 10mM EDTA, 0.5% (w/v) PVP, 50 mM NaCl and 1% (v/v)Triton X. The samples were centrifuged at 4,000 rpm at 4°C for 10 minutes in 50 ml falcon tubes to obtain the soluble protein extract. Supernatants containing protein were moved to screw cap microcentrifuge tubes and concentrated HCl was added to a final concentration of 6N prior to hydrolysis in a heat block overnight at 100°C. The resulting hydrolysed protein samples were centrifuged to remove any remaining debris. The volumes required to contain sufficient amounts of amino acids to enable measurement of their mass isotopomer distributions were tested by drying a range of volumes from one of the original extract, derivatising the test samples with tBDMS and running GC-MS analysis to identify an appropriate volume to use for the rest of the samples. The pH of the aliquot of each hydrolysed sample used for analysis was adjusted by adding 1 ml water to the dried aliquot and neutralising the acidic pH with 1M KOH. The neutralised samples were then dried again and derivatised with tBDMS prior to GC-MS analysis.

2.5.2 Extraction for total protein fractions used for GFP purifications

Plant tissue harvested from seedling cultures at 14 days of growth was patted dry with tissue and ground into powder in liquid nitrogen. The powder was added to homogenisation buffer containing 50 mM HEPES-KOH (pH 7.5), 5% (v/v) glycerol, 10mM EDTA, 0.5% (w/v) PVP, 50 mM NaCl, 1% (v/v) Triton X and Complete protease inhibitor. The Complete protease inhibitor mixture was prepared as instructed by the manufacturer. The weight of homogenisation buffer used was approximately twice that of the tissue. The resulting protein extract was centrifuged at 4,000 rpm at

4°C for 10 minutes, after which the supernatant was collected and centrifuged again at 14,000 rpm at 4°C for 3 minutes in a microcentrifuge tube. The clear supernatant was then moved to a new microcentrifuge tube containing the GFP nanotrap resin.

2.6 Isolation and analysis of protein, soluble metabolites and exudates of root hairs

2.6.1 Isolation of root hairs

Seedling cultures were grown for 14 days in Johnson's medium, after which root hairs were harvested. The fresh tissue was stirred in liquid nitrogen for 20 minutes using a glass rod and then sieved with a metal sieve to separate root hairs from debris that might also have separated from the roots. The container was then rinsed with liquid nitrogen which was subsequently also passed through the sieve to increase the number of root hairs recovered.

2.6.2 Extraction of soluble metabolites and protein from the root hairs

The isolated root hair sample was transferred to a microcentrifuge tube and ground into powder. 1 ml of 100% methanol was added to the sample tube. The samples were centrifuged at 14,000 rpm for 5 minutes to obtain a clear methanol extract for the analysis of soluble metabolites. The entire methanol-soluble extract was dried in a speed-vac and derivatised with MSTFA for GC-MS analysis. The remaining insoluble material was transferred to a microcentrifuge tube and supplemented with 500 µl 6N HCl prior to hydrolysis in a heat block overnight at 100°C. After hydrolysis the samples were centrifuged for 5 minutes at 14,000 rpm to remove any solid material and debris. The samples were then dried and derivatised as described for the whole root total protein fraction (Section 2.5.1).

2.6.3 GC-MS sample preparation for exudates and soluble metabolites

To analyse amino acids and sugars in the soluble fraction and the metabolites in the exudate, samples were derivatised with MSTFA. Dry samples were dissolved in 40 µl of methoxyaminhydrochloride (20 mg/ml in pyridine) and then incubated for two hours at 37 °C and 950 rpm. The samples were centrifuged briefly to recover and condensate from the lid and 70 µl of MSTFA reagent was added

followed by incubation at 37 °C and 950 rpm for 30 minutes. The samples were transferred to glass crimp-top sample vials, sealed and analysed with GC-MS.

2.6.4 Identification and quantification of metabolites from the exudate

The GC-MS spectra from analysis of exudate samples were baseline corrected using the MetAlign software. After this the AMDIS software connected to the NIST metabolite database was used to identify the peaks in the spectra. Using the retention times and masses identified in AMDIS, Agilent Chemstation was used to quantify the peaks based on the peak area relative to that of the internal standard ribitol with a known concentration, taking into account the fresh weights of cultures and the proportion of the sample extract used in the analysis.

2.7 Expression and purification of the GBP antibody in Escherichia coli

2.7.1 Plasmid purification and separation of plasmid DNA on gel

2 µl of a plasmid preparation containing the pHEN6 plasmid with the coding sequence for GFP binding protein (pHEN6-GBP) was used to transform 50 µl DH5α competent cells with heat shock. Details of the complementarity-determining regions and modifications on the nanobodies have been described by Saerens and group (Saerens et al., 2008; Saerens et al., 2005).The pHEN6-plasmid is a phage display plasmid used for the expression of the llama single chain antibody (Conrath et al., 2001). It is based on the pHEN4 and pHEN1 plasmids (Arbabi Ghahroudi et al., 1997), with the cloning site for the camel V_HH gene (introduced in pHEN4) inserted and a hemagluttinin tag and geneIII replaced by a His_6-tag. The pHEN4 phagemid is derived from pUC with an F1 origin of replication. The nanobody expression is under a *Plac* promoter and induced with IPTG. The expressed protein is directed to the periplasm with a pelB leader signal sequence. Like pHEN4, the pHEN6 plasmid contains an ampicillin resistance gene.

The cells were incubated with the plasmid for 10 minutes on ice, after which they were heat shocked at 42 °C for 60 seconds. The cells were then allowed to recover on ice for 10 minutes before addition of 900 µl of Super Optimal broth with Catabolite repression (SOC) medium. SOC is prepared by addition of 2% w/v tryptone, 05% w/v yeast extract, 8.56 mM NaCl, 2.5 mM KCl and 10 mM $MgCl_2$,

10 mM $MgSO_4$ and 20 mM glucose to 1000 ml of deionised water. Samples were incubated at 37 °C with shaking for 60 minutes, after which the cells were centrifuged and resuspended in 100 µl of LB medium. Half of this volume was spread on LB plates that contained 100 µg/ml ampicillin and the plates were incubated at 37 °C overnight.

Antibiotic resistant single colonies were used to start cultures for plasmid purification, which was done with the Qiagen plasmid purification kit following manufacturer's instructions. The plasmid was run on agarose gel to confirm its size and its identity was checked by restriction fragment analysis after digestion with Nco1 and Not1 to release the coding sequence for GFP binding protein. Another fraction of the purified plasmid was recovered from the agarose gel after electrophoresis using a Qiagen gel extraction kit following manufacturer's instructions. This pure plasmid fraction was used to transform the expression strain. Agarose gels were made by dissolving 0.5 g agarose in 50 ml 1xTris-Acetate-EDTA (TAE) buffer with 1 µl of 10 mg/ml ethidium bromide stock solution and fractionation was performed at 80V in an electrophoresis unit.

2.7.2 Protein expression and purification

BL21 competent cells were transformed by the heat shock procedure as described for plasmid cloning (Section 2.7.1) using 2 µl of plasmid preparation (pHEN6-GBP) in 50 µl competent cells. The only difference was that 2% (w/v) glucose was added to all medium and plates.

Two types of media were used for growing bacteria. Luria Bertani (LB) medium was prepared by dissolving 10g of tryptone, 10 g of NaCl and 5 g yeast extract in 950 ml of deionised water with shaking. The pH was adjusted to 7 with 5 N NaOH and the final volume was adjusted to 1 litre. The medium was autoclaved for 20 minutes at 15 psi. Terrific broth (TB) was prepared by dissolving 12 g of tryptone and 24 g of yeast extract into 900 ml of deionised water and shaking. Once the solutes had dissolved, 4 ml of glycerol was added. In a separate bottle, 100 ml of solution containing 0.17 M KH_2PO_4 and 0.72 M of K_2HPO_4 was prepared. Both solutions were sterilised by autoclaving for 20 minutes at 15 psi and mixed under sterile conditions after autoclaving.

Single colonies were picked from the plates to inoculate starter cultures consisting of 10 ml of LB medium supplemented with 2% (w/v) glucose and 100 μg/ml ampicillin, and these were grown overnight. A larger culture of 50 ml of LB medium was then started by adding 2.5 ml of the overnight culture to fresh LB medium with 2% (w/v) glucose and 100 μg/ml ampicillin. Ampicillin was added to 450 ml of TB medium to a final concentration of 100 μg/ml. The culture was inoculated using 50 ml of the overnight culture and grown at 37 °C to an OD of 0.8-0.9 before induction with 1mM IPTG. At this point the temperature was lowered to 26°C and induced expression cultures were grown overnight.

Osmotic shock as described in the Qiagen protocols for His-tag purification: cells were resuspended in buffer (30mM Tris-Cl, 20% sucrose (w/v), pH 8) and kept on ice while slowly adding 500 mM EDTA before incubating the cell for a further 10 minutes on ice with gentle stirring. Cells were collected by centrifugation and the supernatant collected for analysis by SDS-PAGE. Ice cold 5 mM $MgSO_4$ was added and cells were incubated on ice for 10 minutes with stirring. The supernatant collected after centrifugation, containing the periplasmic fraction, was analysed by SDS-PAGE to confirm the presence of the GFP-binding protein. The protein was then purified from the periplasmic fraction with Ni-NTA resin (Qiagen) following manufacturer's instructions using imidazole elution buffer. The wash buffer consisted of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole adjusted to pH 8 using NaOH. The elution buffer contained 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole adjusted to pH 8.0 using NaOH. The samples were bound to the resin for one hour or longer with agitation at 4 °C. The resin was washed two times with wash buffer before elution with the imidazole elution buffer. The protein concentration of purified samples was tested with Bradford's assay using bovine serum albumin (BSA) as a standard.

2.7.3 SDS-Polyacrylamide Gel Electophoresis (SDS-PAGE)

SDS-PAGE was performed on gels containing 12% acrylamide and were run at 80V. Dye solution was added to samples (final concentrations in sample were 0.002% (w/v) bromophenol blue, 0.7135 M β-mercaptoethanol, 10% (v/v) glycerol, 2.5 % (w/v) sodium dodecyl sulphate (SDS), 1M Tris/HCl

pH 6.8) and samples were incubated at 90 °C for 10 minutes prior to transfer to wells. Gels were stained with Coomassie Brilliant Blue (1 g/l in 50% (v/v) methanol, 10% (v/v) acetic acid and 40% H_2O).

2.8 Immunoaffinity purification of GFP from plant protein extracts

Nanotrap beads were formed by coupling 1.5 mg of GFP binding protein to 1 ml of N-hydroxysuccinimide-Sepharose following the manufacturer's instructions and with an overnight incubation of the protein with the Sepharose beads in a buffer consisting of 0.2 M $NaHCO_3$, 0.5 M NaCl, pH 8.3. After blocking any non-reacted groups on the resin with 0.5 M NaCl, pH 8.3 for two hours and washing the medium with the recommended method of two alternating buffers (0.1 M Tris-HCl pH 8.5 and 0.1 M acetate buffer, 0.5 M NaCl pH 5) the medium was ready for use. This resin was used for the highly specific purification of GFP from plant extracts. To purify GFP, 50 µl of the anti-GFP nanotrap resin was added to clear total protein extract in a microcentrifuge tube and incubated for 30 minutes. The supernatant was collected by centrifugation and the resin was washed five times with 1 ml of the dilution buffer containing 300 mM NaCl (Rothbauer et al., 2008). The bound GFP was then eluted with 50 µl of citrate buffer (pH 3.2).

2.9 Confocal imaging and fluorescence measurements

2.9.1 Confocal imaging

Whole roots of 14-day old seedlings with constitutive or cell type specific expression of GFP grown in hydroponic cultures were taken directly from culture, placed on glass plates and mounted under a coverslip with distilled water. A Zeiss LSM510 confocal microscope was used for the confocal imaging. Propidium iodide (PI) was used in most experiments to stain the cells to show the location of GFP expression in the roots. Roots were dipped in 10 mg/ml solution of PI immediately prior to imaging. After staining, the roots were washed with water prior to being mounted under a coverslip in distilled water.

The confocal microscope was equipped with lasers for 458- and 543-nm excitation and images were collected with a Zeiss 25x 0.8 Plan-Neofluar multi-immersion lens in multitrack mode. The 475-525

nm and 585 nm emission filters were used to record the images of GFP expression in roots of the seedlings. The ratio of the 458 nm/543 nm laser power was kept constant at 1:1. Pixel time was set as 1.6 µs. Pinhole settings were 64 µm for the filter BP 475-525 and 62 µm for filter LP 585 (channels 2 and 3 respectively). The imaging parameters were kept constant during the measurements. The instrument is based at the confocal imaging facility at the Department of Plant Sciences, University of Oxford.

2.9.2 Fluorescence measurements

GFP expressed in *Escherichia coli* was purified with affinity purification (His-tag) as described previously for the purification of GBP protein (2.7.2) and the protein concentration was determined with a Bradford assay using BSA as a standard. A dilution series of the standard was prepared on a black 96-well microplate. Samples of GFP purified with the GFP nanotrap were included on the same microplate and the total volume of all samples was adjusted to 100 µl with deionised water. Measurements were taken using a Beckman Coulter DTX880 microplate reader (excitation wavelength 485 nm, emission wavelength 535 nm). The reference readings obtained for known concentrations amounts of GFP purified from bacteria were used to estimate the quantities of GFP obtained from plant extracts.

Alternatively, a commercial 96-well microplate where the wells had been coated with GFP binding protein was used with the same instrument and methods as above, but instead of adjusting volumes with water, the elution volume recommended by the manufacturer was used.

2.10 Purification of GFP samples for GC-MS

2.10.1 Ultrafiltration

Amicon ultra 0.5 ml spin columns containing an ultrafiltration membrane with a 10 kD exclusion limit were used to desalt the eluted GFP samples, and especially to remove citrate introduced in the nanotrap elution buffer, prior to subsequent analysis. The sample was transferred to an individual spin column and the spin column was filled with deionised water (450 µl). The water was allowed to flow through with centrifugation and the column was filled four more times with water and the

centrifugation repeated. At the final step the samples were centrifuged until the total volume reached the minimum of 20-50 µl. The samples were then collected from the column into a new microcentrifuge tube.

2.10.2 Solid phase extraction

Solid phase extraction was done to immunopurified GFP prior to hydrolysis on a 96-well microplate using C_{18} ziptips with a 0.6 µl resin bed and a multichannel P10 pipette. The extraction was done following manufacturer's instructions: the washing solution consisted of 0.1% (v/v) trifluoroacetic acid (TFA) in water, elution solution was 60% (v/v) acetonitrile in water with 0.1% TFA. The tips were washed three times with acetonitrile, followed by three washes with elution solution and finally three washes with 0.1% TFA. The samples were bound by separately aspirating and dispensing each 10 µl volume of total sample through the filter six times. The filter was washed three times with washing solution before elution with 5 µl of the 60% acetonitrile solution. The resulting samples in acetonitrile solution were dried and subsequently resuspended in 50 µl of deionised water. These samples were then combined, if necessary, for further analysis.

2.11 Gas chromatography – mass spectrometry

2.11.1 GC-MS sample preparation for amino acid samples derived from total protein

Amino acids were derivatised by adding 25 µl of pyridine to dry samples and incubating with 950 rpm shaking at 37 °C for 30 minutes followed by addition of 35 µl of the tBDMS derivatisation reagent and incubation with 950 rpm shaking at 65 °C for 30 minutes. The samples were then moved to glass GC-MS vials and sealed.

2.11.2 GC-MS sample preparation for samples containing small amounts of GFP

After immunopurification and filtering steps, GFP was hydrolysed in screw cap microcentrifuge tubes by addition of 50 µl of 6N HCl and overnight incubation at 100 °C. If samples from several seedling cultures were used, they were combined in the same microcentrifuge tube. The samples were dried and derivatised by addition of 8 µl of pyridine and incubation with 950 rpm shaking at

37 °C for 30 minutes followed by addition of 12 µl of tBDMS reagent and incubation with 950 rpm shaking at 65 °C for 30 minutes. The samples were then moved to glass GC-MS vials and sealed.

2.11.3 GC-MS analysis

Standard GC-MS runs were performed with an Agilent 7890A GC coupled to a 5975C quadrupole MS detector and 7683B series injector with electron impact ionisation (70 eV). The column used was Varian VF5 ms column (30 m, 10 m guard column, 0.25 mm inner diameter). Before injection, the GC-MS was calibrated by using the Agilent AutoTune method. For analysis of amino acids derivatised with tBDMS, the injection (1 µl) was performed in splitless mode at 230 °C and the oven temperature programme was set to increase from 120 °C (for 5 minutes) to 270 °C (for 4 minutes) at a rate of 4 °C per minute, then to 310 °C (for 10 minutes) at a rate of 20 °C per minute with a continuous carrier gas (He) flow of 1.3 ml/min. The scan speed was 4.38 scans/second.

For soluble metabolites and metabolites from the root hair exudates, the injection (1 µl) was performed at 230 °C and the oven temperature was initially held constant at 70 °C for five minutes, then ramped up to 350 °C at 5°C per minute, immediately decreased to 330 °C (five minutes) at 120 °C per minute and decreased to 70 °C at 120 °C per minute. The carrier gas flow (He) was 0.6 ml/min.

Mass spectra were acquired for m/z 146-600. Solvent delay of 10 minutes was included with the mass spectrometer turned on after elution of the solvent from the column. The data was acquired in centroid mode using the Agilent Chemstation software. The instrument used was based at a mass spectrometry facility at the Department of Plant Sciences, University of Oxford.

Trace GC Ultra from Thermo Fisher coupled to an ISQ single quadrupole mass spectrometer was used to test larger volume injections. The instrument used was based at Technische Universität Dortmund, Institut für Umweltforschung der Fakultät für Chemie und Chemische Biologie (INFU) and parameters not specified here followed the standard protocols used at the facility. Scan time was 0.2 s and injection volume was set as 5 µl. Solvent delay of 9 minutes was included and the temperature programme remained the same as that described for the Agilent 7890A GC coupled

system described above with the same total running time. As in previous experiments, electron impact ionisation was used. The instrument was calibrated using the AutoTune_EI method immediately before running the samples. Data was acquired in centroid mode using Xcalibur software.

An Agilent 7890B GC coupled to 7200 Q-TOF was used for measurements where noted. This instrument was available at the Mass Spectrometry Research Facility, Department of Chemistry, University of Oxford. The method used with the Agilent 7200 was based on that used with Agilent 5795C and injection volumes were 0.5-1µl, depending on the experiment. The temperature programme remained the same as that previously described for the Agilent 7890A GC coupled system and solvent delay of 9 minutes was included. Electron impact ionisation was used with the quadrupole time of flight analyser coupled to an electron multiplier detector. The instrument has 4GHz ADC electronics, which enables sampling rate of 32 Gbits/s and the sampling frequency was 0.5. Data was automatically acquired in centroid mode for mass range 40 to 1700 m/z using Agilent MassHunter software, but only the range 40 to 600 m/z was saved.

2.11.4 Processing of mass isotopomer data

Raw data from spectra were processed into mass isotopomer data by two main correction steps: baseline correction with MetAlign software in order to take account baseline noise, and mass correction with MSCorr program run on MATLAB to correct for the contribution from atoms in the functional groups added during derivatisation as well as atoms other than carbon to the mass isotopomer profile of the molecule

The baseline correction procedure was run using standard parameters routinely used for data used for metabolic flux analysis experiments in the laboratory: mass bin resolution was set to 0.65, retention begin and end scan numbers were set as 1 and 10,000 respectively, peak slope factor was set as 1, peak threshold factor as 2, peak threshold absolute value as 0 and average peak width at half height as 5. Settings for initial peak search criteria, scaling options and tuning alignment options and

criteria were kept in their default settings. Peak selection significance percentage was set as 99, minimum ratio between means and minimum S/N ratio as 2.

The full details of the MetAlign software have been described by the developers of the software (Lommen, 2009) and in several published papers where metabolite data from LC-MS and GC-MS experiments have been corrected using the software. While the software is capable of data alignment, the main function used here was to correct for baseline noise associated with the normal running of a mass spectrometer attached to a chromatography system, such as column bleed and chemical and detector noise. The algorithms allow the detection of possible significant measurement peaks based on user defined minimum thresholds and a local noise value as function of time. An estimated noise matrix is used while the algorithm assesses the data in the file for an increase above the threshold from left to right or right to left. Increases above the threshold are noted and combined, if they are next to each other, to form peak regions. Anything outside the peak regions is considered as a baseline signal. The baseline regions are set to zero before repeating the procedure a second time using noise values that are half of the original values. Finally, any small peaks that fall below a user defined absolute threshold are removed.

The main concern with use of any algorithm for processing of data is its effect on interpretation of the results. In the case of MetAlign, the processing helps to discriminate between signals derived from chemical fragments and the background noise. If signal intensities are low enough to be considered to form part of a baseline region rather than a peak region, arguably the measurement is not very reliable in the first place. In the case of overloading of samples, baseline correction applies an artificial maximum value on the peaks that cannot be digitized by the detector. In the vast majority of experiments described in this thesis the analytes of interest (amino acid peaks) are measured at considerably lower ion counts than other peaks arising through derivatisation. These higher peaks are likely to reach saturating levels before the analytes of interest. Therefore it is unlikely the analytes of interest would be present at levels where the peaks could not be digitized by the detector, which would lead to correction of data.

Data on individual metabolites from the baseline corrected spectra were collected using Agilent Chemstation and the data obtained from the spectra were saved in a text file format for subsequent processing using MSCorr. The mass correction process takes into account the naturally occurring stable mass isotopes in the fragment by use of infinite matrix calculus and assesses the consistency of the measurements (Wahl et al., 2004). The corrected data were in many cases used to calculate the average percentage of ^{13}C in a fragment according to the following relationship:

$$FL = \frac{\sum_{i=0}^{n} i \cdot m}{n \cdot \sum_{i=0}^{n} m_i}$$

where FL stands for fractional labelling, n represents the number of carbon atoms in the carbon skeleton of the amino acid fragment, i indicates the different mass isotopomers and m_0 is the corrected fractional abundance of monioisotopic molecules containing only ^{12}C and $m_i>0$ corresponds to abundances of fragments with heavier masses (Xiong et al., 2010). An example of the process is shown in Table 1.

Table 1: **Calculation of average ^{13}C abundance from data generated by MSCorr from total ion counts in a fragment with a carbon skeleton consisting of four carbons.**

Mass peak	Value (MSCorr)	^{13}C substitutions	Carbon substitutions x value	Σ divided by number of carbons (4)	Percentage
M+0	0.575	0	0 x 0.575		
M+1	0.1664	1	1 x 1.664		
M+2	0.2063	2	2 x 0.2063		
M+3	0.0314	3	3 x 0.0314		
M+4	0.0209	4	4 x 0.0209		
Total				**0.1892**	**18.92%**

2.12 Metabolic modelling

All metabolic modelling was performed using the 13CFLUX software platform (Wiechert, 2001). The steady-state isotopomer composition of metabolites produced by a defined set of fluxes in a model network supplied with a substrate of specified isotopic composition was determined using the CumoNet module. Conversely, fluxes that best described the measured (or simulated) redistribution of label through a modelled metabolic network were fitted by non-linear optimization using the sequential quadratic programing method, Donlp2 (developed by P. Spellucci, http://www.mathematik.tu-darmstadt.de/fbereiche/numerik/staff/spellucci/DONLP2/donlp2doc.pdf), based on iterative numerical fitting procedures. Typically, Donlp2 was employed in 13CFLUX to perform Monte Carlo simulations (implemented using the –mc argument) to derive 100 optimised flux solutions based on a random set of initial flux values and a series of isotopomer values obtained by statistical sampling from the population defined by the measured (simulated) measurement and its

corresponding SE. The means, residua and upper and lower quartile values were collated for the estimated flux solutions. The residua (residual sum-of-squared difference between predicted and measured (simulated) values) of flux solutions that were extracted from the large output files generated by 13CFLUX using the "grep" command. Individual flux solutions based on the residuum values were then obtained from the same output files using the Feasible residuum extractor, a script tool written by Shyam Masakapalli (University of Oxford) The means, residua and upper and lower quartile were collated for the estimated flux solutions for each set of measured (simulated) isotopomer values.

3 *In silico* analysis of flux determination in heterogeneous systems

Chapter 1 introduced the background to cell type specific metabolic flux analysis, considered the different directions the work might take, and discussed the importance of developing new strategies to access the metabolic information in specific types of cells. This chapter aims to investigate whether current standard methods, which are typically based on measurements of metabolites extracted from a combination of different cell types (and thus providing a weighted average of the labelling patterns of metabolites across the tissue), provide an accurate indication of the average fluxes from heterogeneous tissue. If average flux maps of systems with multiple flux phenotypes are not accurate, cell type specific strategies are required not only to study the metabolism of the different cell types, but in order to provide any realistic assessment of fluxes in differentiated tissues.

Simulated isotopomer data obtained using *in silico* models of simple networks consisting of only a few fluxes were mixed in different ratios to produce composite data sets representing different contributions of the two component flux networks to the label distribution in the metabolites within the overall system. The global best fit fluxes obtained from Monte Carlo simulations using the composite measurements were then compared against the weighted averages of fluxes in the models used to create the simulated data set. These studies were designed to examine how reliable flux maps of heterogeneous tissue are without development of new strategies for cell type specific metabolic flux analysis. This chapter presents a comparison of how well flux estimates obtained from composite datasets report on the actual fluxes in the system using a range of small scale models that differ in the number and type of free flux, thereby represents the variation that can exist within a metabolic network.

3.1 Introduction

The vast majority of plant metabolic flux analysis studies reported to date have been carried out in cell cultures or other systems generally anticipated to consist (either entirely or predominantly) of a single homogeneous cell type. All such systems are assumed to consist of cells with a single flux phenotype and the assumption is supported by the production of flux maps with good fits from the labelling data (Kruger and Ratcliffe, 2009). Without specific strategies for obtaining metabolites

such as protein from different cells, studies of differentiated tissues in cells consisting of more than one cell type would have to rely on combined metabolite pools extracted from a mixture of cells to measure stable isotope labelling. In the few examples where heterogeneous tissues have been studied with metabolic flux analysis, the possible issues with studying differentiated systems have been largely ignored and average data has been used to report on the fluxes (Niklas et al., 2010). It is therefore important to understand how well such average flux maps (derived from the composite or aggregate metabolite labelling data) report on the actual fluxes in the metabolic network.

The aim of the chapter was to evaluate whether flux maps of heterogeneous tissues are accurate representations of flux phenotypes. The specific objectives were to use small models of very simple networks of fluxes to test with different combinations of free net and exchange fluxes how the results of Monte Carlo simulations are affected by use of aggregate datasets from combinations of simulated data. The global best fit solutions obtained for each small scale network were compared against the average fluxes that were used to simulate the datasets. The comparison established that there are differences between the average fluxes in the system and the global best fit solutions for some of the small scale models depending on the network. The nonlinear relationship between isotopomer distributions and the deduced fluxes was identified as the cause of the differences.

3.1.1 Metabolic modelling using simple networks and Monte Carlo simulations to obtain global best fit solutions

Metabolic networks consist of fluxes, which in a model used for metabolic flux analysis can be either determined, constrained or free. Reactions in the metabolic network are also potentially reversible: normally they are considered as forward and reverse fluxes, but they can be redefined as net and exchange fluxes. Due to the intrinsic stoichiometry constraints of the network, the flux distribution through the network can be fully described by the values of a subset of fluxes. Constrained fluxes are limited by thermodynamic considerations or independent measurements. The free fluxes, which form the remainder of the defining subset of the fluxes can be varied during the optimisation procedure that fits the experimental data to the metabolic network. The remaining fluxes are the

dependent fluxes, which are determined by reaction stoichiometry and carbon transitions described for the metabolic network as well as the values chosen for constrained and free fluxes.

When the inputs and outputs of a system remain constant, a specific reaction network is characterised by the values of the free fluxes. Due to the non-linear relationship between fluxes and isotopomer composition deducing the free fluxes requires a robust approach for assessing the reliability of the deduced flux distribution. This is provided by repeated Monte Carlo simulations that allow the re-optimisation of the best fit solution to obtain the global best fit characterised by the low residuum of the solutions (Kruger et al., 2012). The best fit solutions reporter in this chapter were obtained using 13CFLUX (Wiechert, 2001).

The major aim of the work carried out here was to investigate the impact of averaged data on the accuracy of the deduced fluxes. The specific objectives were to find out: 1) do the composite data allow accurate determination of fluxes, 2) is the accuracy of the flux estimates determined from such data influenced by the actual flux values, and 3) does the structure of the network and the variation in the number and distribution of free fluxes influence the accuracy of the determined fluxes?

The simplified networks used here consist of different combinations of net and exchange fluxes to elucidate how modelling with average data affects metabolic networks. Each of the networks was explored by systematically varying one of the fluxes in the system while keeping the other fluxes constrained. For each of the simulations, 100 Monte Carlo simulations were run and the resulting flux estimates were compared to the expected flux that was based on the contribution of the two isotopomer datasets used to create the average dataset.

3.2 Results

3.2.1 Small scale flux networks containing a single free flux yield accurate estimates of the aggregate flux

The simplest examples of networks with free fluxes would contain either one free net flux or one free exchange flux. In the first toy model, an uptake flux R0 goes from substrate S to metabolite A (Figure 4). The uptake flux (net flux into the system) of the two carbon compound consisting of

atoms 1 and 2, indicating position in the molecule, is constrained to one in all of the models used for this work and has no exchange flux (Table 2). Flux then goes from A to metabolite B through R1, which does have an exchange flux. In the first model the goal is to predict flux distribution when the exchange flux R1_xch is set as a free flux back to the two carbon metabolite A. From B the flux goes via R2 out of the network. R2 is set as determined and the exchange flux back into the system is constrained to 0. These were used as the starting conditions for the first simple model, with other fluxes in the network set as determined or alternatively constrained to zero for exchange fluxes for the Monte Carlo simulations. The network used is shown in Figure 4.

Initially, the CumoNet function in13CFLUX was used to determine the isotopomer composition of each of the metabolites in the network under conditions when the exchange flux through R1 was either 0, 0.5 or 0.99. The simulated data was combined in different ratios by calculating weighted means of the different isotopomer datasets. For the exchange flux, datasets were created by combining the fractional abundance of isotopomers obtained when R1_xch=0.99 and 0 or when R1_xch=0.50 and R1_xch=0 (each in ratios 3:1, 2:1, 1:1, 1:2, 1:3). The flux estimates obtained from running 100 Monte Carlo simulations with each of these combined datasets were compared to the aggregate flux derived from the fluxes in the two original networks and the relative contributions of the original isotopomer abundances to the combined dataset. The results establish that in this network 1) fluxes are precisely defined by the combination of isotopomer data used for metabolic flux analysis, and 2) the combined dataset accurately reflects the aggregate flux through the simulated heterogeneous network irrespective of the relative fluxes through the component systems or their relative contributions to the combined dataset (Figure 4).

A simple model where a single net flux could be varied instead of the exchange flux was also constructed using largely the same constraints (Table 3). Here the exchange flux R1_xch was constrained to 0 and the net flux via R1 goes to separate pools of metabolites Ba and Bb (Figure 5). Another net flux, R2, was added on the side to allow R1 to vary and output fluxes from Ba and Bb were added (R3 and R4). Their exchange fluxes were constrained to 0.

For the R1 net flux, datasets were created by combining the fractional abundance of isotopomers obtained when R1_net = 0.99 and 0 or when R1_net = 0.50 and R1_net = 0 (each in ratios 3:1, 2:1, 1:1, 1:2, 1:3). The flux estimates obtained from running 100 Monte Carlo simulations with each of these combined datasets were compared to the aggregate flux derived from the fluxes in the two original networks and the relative contributions of the original isotopomer abundances to the combined datasets. The results were similar to the first network 1) fluxes are precisely defined by the combination of isotopomer data used for metabolic flux analysis, and 2) the combined dataset accurately reflects the aggregate flux through the simulated network irrespective of the relative fluxes through the component systems or their relative contributions to the combined dataset (Figure 5).

Figure 4: **Analysis of deduced fluxes in a heterogeneous network containing a single variable exchange flux**. The small scale network used to simulate data with one free exchange flux and the linear regression analysis between the expected flux distribution used to simulate the datasets and the global best fit solution. The line indicates equivalence between the flux solution and expected flux based on contributions of different datasets. Standard deviations (SD) obtained from Monte Carlo simulations are smaller than the area of the symbol. Flux estimates deduced from mass isotopomer data from individual datasets reporting on a single network of fluxes are indicated by red symbols and flux estimates deduced from combined datasets are indicated by black symbols.

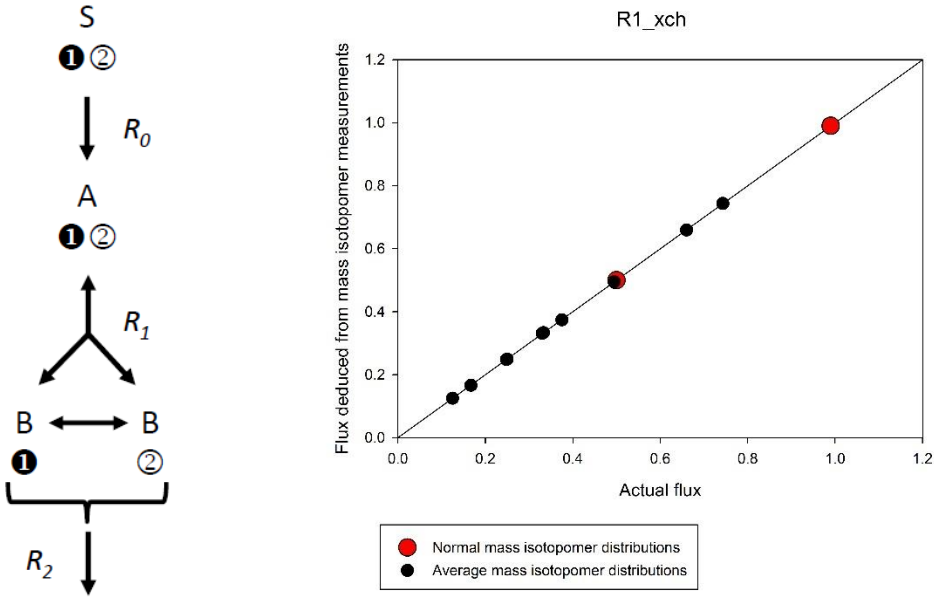


Table 2: **Flux assignments for the toy model used in Figure 4.**

Flux name	***Free / Constrained / Determined***	***Net***	***Free / Constrained / Determined***	***Exchange***
R0	C	1	C	0
R1	D		F	Variable
R2	D		C	0

Figure 5: **Analysis of deduced fluxes in a network containing a single variable net flux.** The small scale network used to simulate data with one free net flux and the linear regression analysis between the expected flux distribution used to simulate the datasets and the global best fit solution. The line indicates equivalence between the flux solution and expected flux based on contributions of different datasets. Standard deviations obtained from Monte Carlo simulations are smaller than the area of the symbol. Flux estimates deduced from mass isotopomer data from individual datasets reporting on a single network of fluxes are indicated by red symbols and flux estimates deduced from combined datasets are indicated by black symbols.

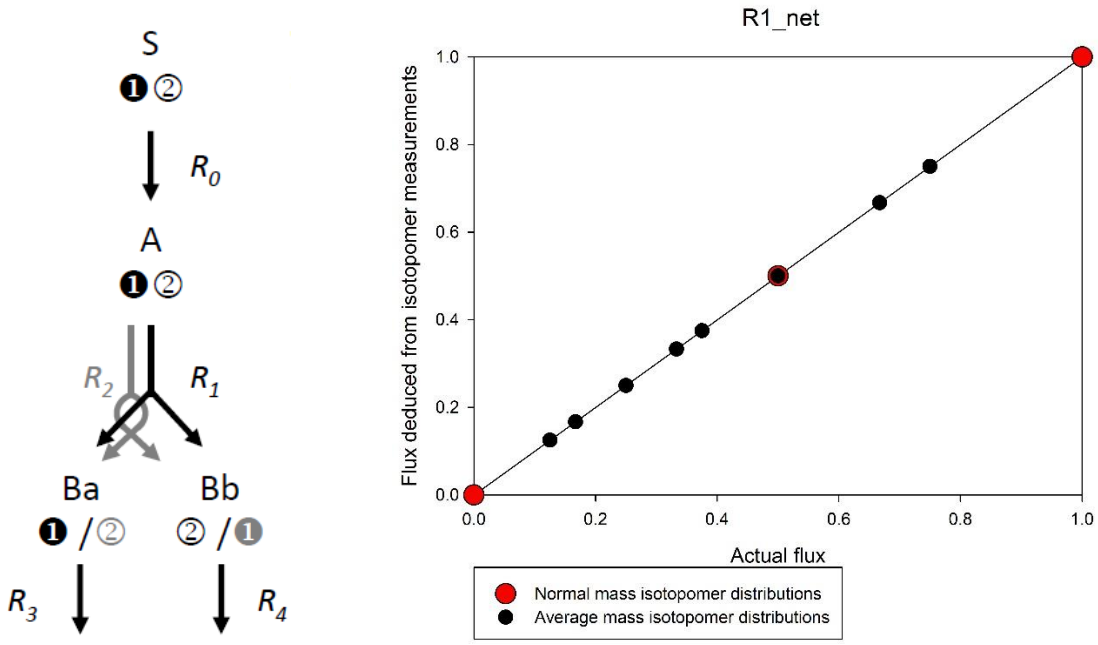


Table 3: **Flux assignments for the toy model shown in Figure 5.**

Flux name	***Free / Constrained / Determined***	***Net***	***Free / Constrained / Determined***	***Exchange***
R0	C	1	C	0
R1	F	Variable	C	0
R2	D		C	0
R3	D		C	0
R4	D		C	0

3.2.2 Aggregate flux estimates derived for small scale flux networks defined by two free fluxes are unreliable.

The small models with one free net or exchange flux represent the simplest case of a metabolic network that can be subjected to this type of analysis. Next the complexity of the networks under investigation was increased by defining the fluxes in the network through two exchange fluxes (R1_xch and R2_xch) or one net flux (R1) and one exchange flux (R1_xch). The weighted means of simulated data were again used to create sets of combined isotopomer data with different relative contributions from the different initial sets of data.

The network for the simple toy model was first modified by allowing an exchange flux for R1 and introducing a flux R2 between Ba and Bb. As a first step the isotopomer composition of metabolites produced by networks with defined flux distributions were calculated. The datasets for each flux were then combined to produce the simulated datasets that would be obtained in a heterogeneous system containing a mixture of the different flux networks. For R1_xch flux, datasets were created by combining the fractional abundance of isotopomers obtained when R1_xch = 0, R1_xch =0.50 or when R1_xch = 0.99. Corresponding datasets were created for when R2_xch = 0, R2_xch = 0.50 or when R2_xch = 0.99. For the final network a net flux R2 was introduced between A and Ba and Bb instead of between the two pools. Here datasets were created for one net flux (datasets were created for when R1_net = 0, R1_net =0.05 or when R1_net = 1) and one exchange flux (datasets were created for when R1_xch = 0, R1_xch = 0.50 or when R1_xch =0.99). The flux estimates obtained from running 100 Monte Carlo simulations with each of these combined datasets were compared to the aggregate flux derived from the fluxes in the two original networks and the relative contributions of the original isotopomer abundances to the combined datasets.

In the case of two free exchange fluxes as well as one free net flux and one free exchange flux, the network fluxes in the original models that were used to generate the simulated datasets were accurately and precisely determined by the isotopomer datasets generated from these models. However, for both of these networks with added complexity, the values of flux estimates differed

from the expected value when the isotopomer data was drawn from two datasets representing different flux distributions (Figure 6 and Figure 7). For the model containing two free exchange fluxes, the effect was most prominent when R1_xch was held constant and R2_xch was varied. The flux estimates for the network in which R1 and R1_xch were varied showed deviations regardless of which of the fluxes was kept constant. There was also significant variation in the flux estimates for the flux that was set free but held constant (Appendix 1). The pattern of deviation for the estimated fluxes was not random, suggesting the estimates depicted the isotopomer distributions provided in the datasets. However, the isotopomer data was not faithfully reporting on the flux distribution.

The failure of average data to report on the aggregate fluxes in a heterogeneous system consisting of two or more networks with different flux distributions arises from the fact that the relationship between isotopomer data and deduced fluxes is not linear. This can be demonstrated by examining the isotopomer distributions simulated for the metabolite A in the network in which R1 and R1_xch were varied (Figure 8). The mathematical basis for why the average data might not report on the aggregate fluxes have also been recently described (Ghosh et al., 2014). While the equations described by Ghosh and colleagues do not directly relate to the ones forming the basis for analysis with 13CFLUX, the basis for non-linearity of the relationship is also well known for these sets of equations: repeated solving of nonlinear equation systems forms the basis of the analysis (Wiechert et al., 1999; Wiechert et al., 2001).

3.3 Discussion

Estimation of fluxes in a metabolic network using Monte Carlo simulations to obtain a global best fit of the measurement data relies on meaningful isotopomer measurements from the biological system. Measurements from mixed metabolite pools could potentially be reliable, if the relationship between the isotopomer distribution measured from a metabolite and the deduced fluxes was linear. However, this is not necessarily the case and leads to data that simply reports on the average fractional abundance of stable isotopes in a mixed metabolite pool, not always on the distribution of fluxes. Since the observed results arise from this flaw in using average measurements to build flux maps, the conclusion holds true for most larger networks, including more realistic metabolic

networks. To test this proposal, the approach developed in this chapter has subsequently been applied to a larger scale model, in which similar discrepancies between the expected fluxes and those deduced from the global best fit solutions obtained using combined isotopomer datasets have also been observed (Appendix 2).

The analysis presented in this chapter can help estimate to which extent use of combined datasets affect the accuracy of deduced fluxes. Figure 8 shows the relationship between the fractional abundance of isotopomers of metabolite A the flux through R1 and R1_xch in the simple model containing only one free exchange flux and one free net flux. Here the labelling of carbons in positions 1 and 2 in the model are indicated using the standard convention of 1 for a ^{13}C in the specific position and 0 for ^{12}C. The metabolite pool A is originally derived from the substrate, which in the model network would be equal to a molecule corresponding to 01. In this network, if the flux from A to Ba and Bb goes entirely via R1 with no flux through R2, all of the molecules in the Ba and Bb metabolite pools will be 0 and 1, respectively. Thus, A will retain the original 01 pattern of label irrespective of whether it is formed directly from S via R0 or from Ba and Bb through R1_xch. This situation is represented on the far right of the x-axis of Figure 8 A. If all the pools contain molecules with the same isotopomer distribution, the exchange flux R1_xch cannot be quantified from the measurement data as there is no change observed as a result of molecules moving across the three metabolite pools. Under all other conditions, the fractional abundance of ^{13}C in A is influence by the rearrangement of label brought about through the combination of the flux from A to Ba and Bb through R2 and the resynthesis of A from Ba and Bb through R1_xch. This effect is most pronounced on the extreme left of the x-axis of Figure 8 A when there is no contribution from R1 to the conversion of A to Ba and Bb. Under these circumstances, the most abundant isotopomer of A still matches the labelling pattern in the network substrate (S), but because the entire flux from A to Ba and Bb is through R2, the highest proportion of 00, 11 and 10 isotopomers are also observed. Moving from left to right along the x-axis, the contribution of R2 to conversion of A to Ba and Bb decreases (and that of R1 net increases) thereby decreasing the proportions of the isotopomers formed through rearrangement of the carbon backbone.

However, because the changes in the isotopomeric composition of A are not linear, it becomes increasingly difficult to estimate the flux distribution from the average dataset in the middle, where flux goes through R1, R1_xch and R2. This gives rise to the imprecise flux estimates observed in the middle of Figure 6 B and Figure 7 A and B. An error arising from the difference between a mass isotopomer value derived from two datasets and the value corresponding to the expected flux estimate in the non-linear distribution of measured mass isotopomer values is introduced into each isotopomer's fractional abundance. As the contribution of a single flux at either extreme gets greater, the ability to predict the flux also improves due to decrease in this error. The same argument holds when R1_xch is varied and R1 net flux is set as 0.5 (Figure 8 B). When there is no exchange flux back to metabolite A, the isotopomer pool consists entirely of molecules with the labelling pattern 01. From left to right the number of other mass isotopomers increases as flux via R1_xch increases. Average data do not accurately report on the fluxes in the middle of the range, but are better at predicting the fluxes at both extremes.

Since the error introduced through averaging data depends on the original datasets, it varies between systems. For simple networks consisting of one net or exchange flux, the average data was able to predict the expected average flux because in these instances the isotopomer distributions were linearly dependent on the value of the free flux. Wiechert and colleagues have described mathematically, how the nonlinear nature of the relationship between fluxes and mass isotopomer distributions relates to a quadratic term in the equations that are used to describe reactions in which a metabolite is formed in a bimolecular reaction step (Wiechert et al., 1999). This explains why the results are different when the exchange flux R2_xch between Ba and Bb (one carbon molecules) metabolite pools is varied and when the R1_xch flux between metabolite pools Ba and Bb and A (two carbon molecule) is varied. R2_xch is not a bimolecular reaction step and in addition the relationship between the educt and product equations is not directly affected by one. When R2_xch is held at a specific value the average data reports faithfully on the expected fluxes, while when the same is done with R1_xch there is significant variation from the expected flux distribution. The exception to this is formed when R2_xch is set as 0 (Figure 6 A). When there is no flux between Ba

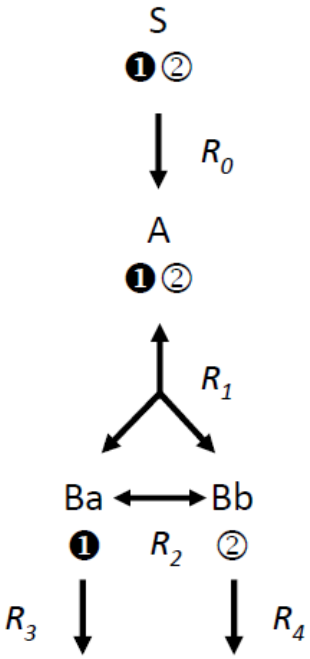
and Bb through R2_xch, there cannot be any fractional abundances different from pool A formed through flux R1_xch, as the two pools each consist of a specific mass isotopomer. Therefore it becomes impossible to quantify R1_xch, since the mass isotopomers that arise are identical to those that make up pool A regardless of the contribution of the exchange flux. In the network with one exchange and one net flux, R1_xch similarly represents a flux between metabolite pools Ba and Bb and the two carbon compound A. The results correspond to having a bimolecular reaction step and inaccurate flux estimates are obtained when both R1 and R1_xch contribute to the flux distribution. However, the simple network described in Figure 4 also contains the reaction step R1_xch and this demonstrates there are networks where it is possible to deduce fluxes from combined mass isotopomer data despite bimolecular reaction steps.

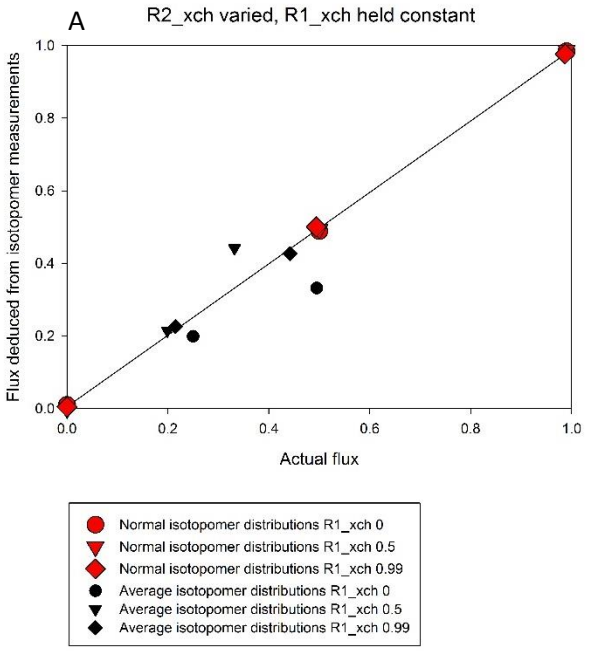
Overall, the differences between the fluxes estimated based on contribution of each dataset to the average data and those obtained as flux solutions were significant, but not extremely large in all experiments where differences were observed. Knowing that the differences can frequently arise in the majority of networks, average flux maps are an inaccurate and unreliable way of presenting fluxes in heterogeneous tissue. However, this is only relevant, if different cell types in the tissue have significantly different metabolic activity. If the metabolism of cell types is very similar, average flux maps may well be able to resolve the aggregate fluxes with sufficient accuracy. Therefore, cell type specific methods to study metabolic activity must be developed in order to better understand what modelling approaches are required and potentially to facilitate cell type specific MFA.

Table 4: Flux assignments for the toy model used in Figure 6.

Flux name	*Free / Constrained / Determined*	*Net*	*Free / Constrained / Determined*	*Exchange*
R0	C	1	C	0
R1	D		F	Variable
R2	D		F	Variable
R3	D		C	0
R4	D		C	0

Figure 6**: Analysis of deduced fluxes in a network containing two variable exchange fluxes.** The small scale network used to simulate data with two free exchange fluxes and the linear regression analysis between the expected flux distribution used to simulate the datasets and the global best fit solution. The line indicates equivalence between the flux solution and expected flux based on contributions of different datasets. Standard deviations obtained from Monte Carlo simulations are smaller than the area of the symbol. Flux estimates deduced from mass isotopomer data from individual datasets reporting on a single network of fluxes are indicated by red symbols and flux estimates deduced from combined datasets are indicated by black symbols. (Next page)


S
❶ ②
R_0
A
❶ ②
R_1
Ba
Bb
❶
R_2
②
R_3
R_4


A
R2_xch varied, R1_xch held constant
Flux deduced from isotopomer measurements
Actual flux
0.0
0.2
0.4
0.6
0.8
1.0
Normal isotopomer distributions R1_xch 0
Normal isotopomer distributions R1_xch 0.5
Normal isotopomer distributions R1_xch 0.99
Average isotopomer distributions R1_xch 0
Average isotopomer distributions R1_xch 0.5
Average isotopomer distributions R1_xch 0.99

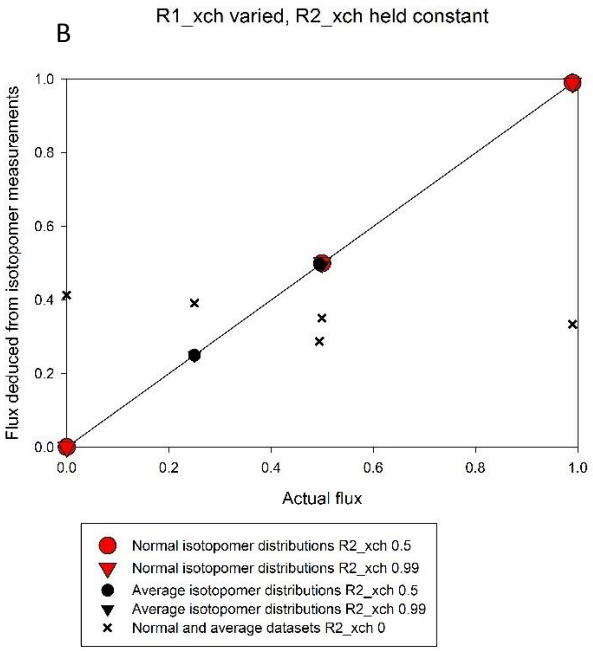
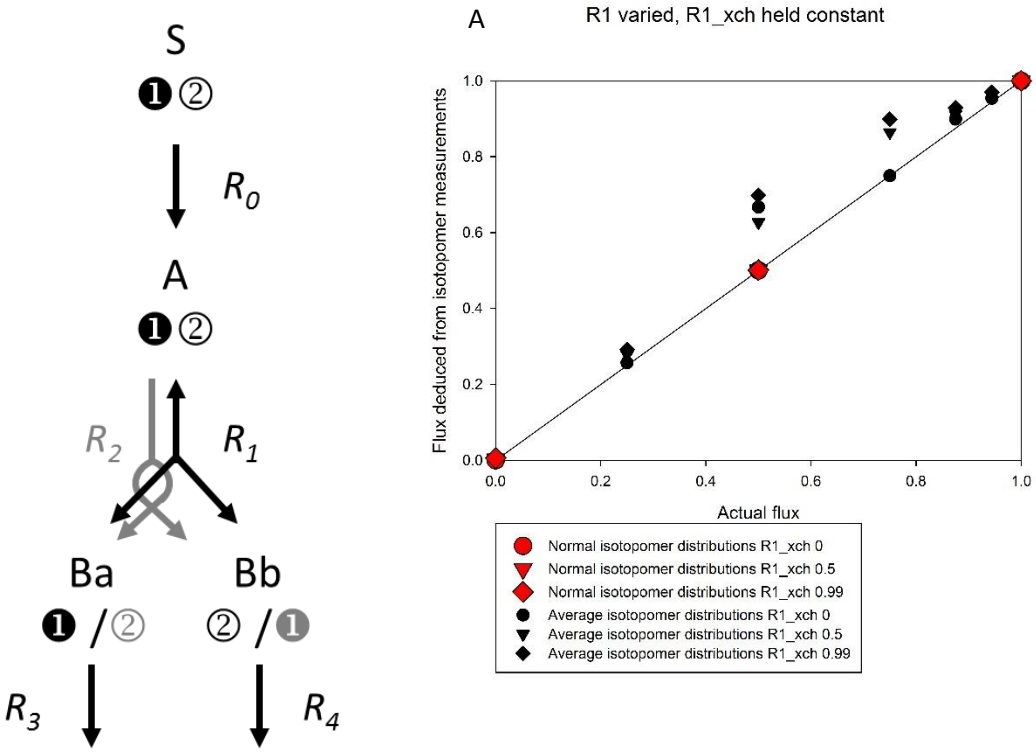

B
R1_xch varied, R2_xch held constant
Flux deduced from isotopomer measurements
Actual flux
0.0
0.2
0.4
0.6
0.8
1.0
Normal isotopomer distributions R2_xch 0.5
Normal isotopomer distributions R2_xch 0.99
Average isotopomer distributions R2_xch 0.5
Average isotopomer distributions R2_xch 0.99
Normal and average datasets R2_xch 0

Figure 7: **Analysis of deduced fluxes in a network containing a single variable net flux and a single variable exchange flux.** The small scale network used to simulate data with one free net flux and one free exchange fluxes and the linear regression analysis between the expected flux distribution used to simulate the datasets and the global best fit solution. The line indicates equivalence between the flux solution and expected flux based on contributions of different datasets. Standard deviations obtained from Monte Carlo simulations are smaller than the area of the symbol. Flux estimates deduced from mass isotopomer data from individual datasets reporting on a single network of fluxes are indicated by red symbols and flux estimates deduced from combined datasets are indicated by black symbols.



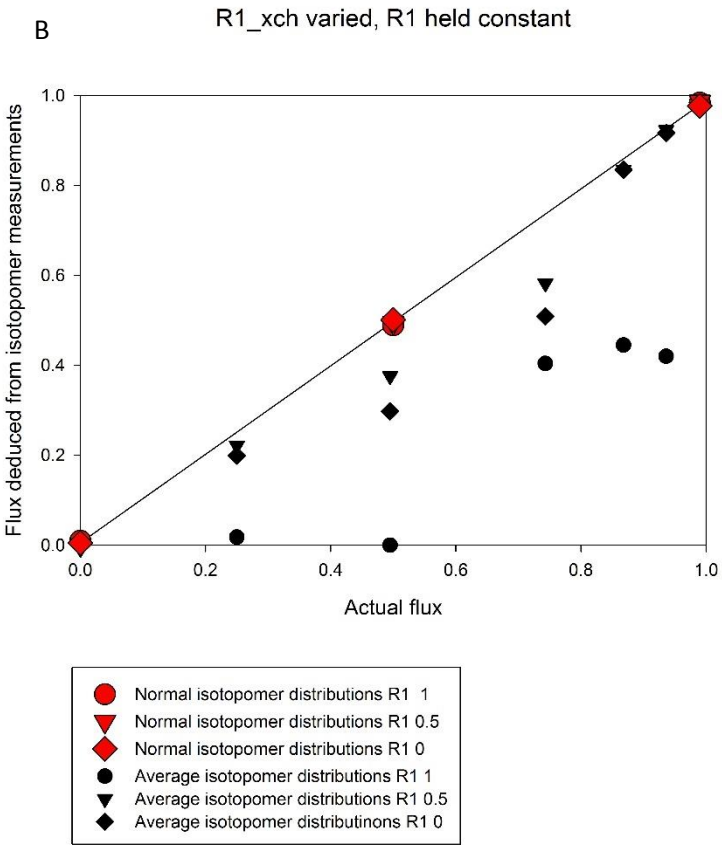
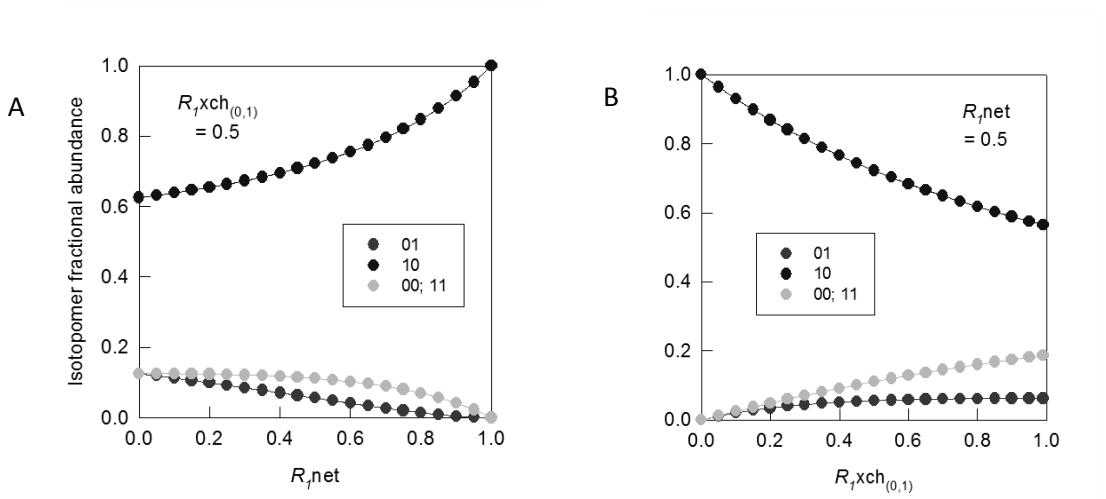


Table 5: Flux assignments for the toy model used in Figure 7

Flux name	***Free / Constrained / Determined***	***Net***	***Free / Constrained / Determined***	***Exchange***
R0	C	1	C	0
R1	F	Variable	F	Variable
R2	D		C	0
R3	D		C	0
R4	D		C	0

Figure 8: **Changes in fractional abundances of isotopomers for metabolite A when fluxes are varied in the network containing a free net flux and a free exchange flux.** Simulations were conducted using the network described in Figure 7. Changes in the isotopomer fractional abundances for Metabolite A when an individual free flux was varied. (A) net flux R1 varied over the range 0-1 with R1_xch held constant at 0.5. (B) R1 normalised exchange flux varied over the range 0-0.99 with R1 net flux R1 is held constant at 0.5. The conventional format for indicating positional labelling in a molecule is used with 0 indicating a non-labelled atom (^{12}C) and 1 indicating substitution with a stable isotope (^{13}C).



4 Development of a protocol for determining the mass isotopomer profile of GFP expressed in Arabidopsis thaliana roots

The *in silico* analyses presented in Chapter 3 demonstrated the need to investigate the metabolic fluxes in different cell types or mixtures of cells using cell type specific methods rather than average data. However, current methods are not capable of reporting on fluxes in different cell types. In order to measure the mass isotopomer distributions in separate pools in different types of cells, a technique is needed that is able to discriminate between labelled metabolites from a sub-group of cells without the risk of causing changes to the metabolic signals during the extraction. As discussed in Chapter 1, one of the main cellular components exploited in metabolic flux analysis is protein. It is composed of amino acids that are synthesised through different biosynthetic pathways from a diverse range of precursors and therefore is particularly well suited for reporting on differences in central carbon metabolism. A reporter protein expressed in only one cell type may be able provide information on the metabolism of those specific cells. As long as cellular protein consists of primarily amino acids produced in that particular cell, it would be possible to derive information on the metabolic activity of the cell from the component amino acids of the reporter protein.

Published work has shown the main challenges for the use of a reporter protein strategy to study sub-sets of cells in mixtures are related to the small amounts of protein that can be purified from the cells and contamination from different sources. The most difficult contamination is the type that affects the detection of the mass isotopomer distributions from the amino acids derived from protein. The aim of this chapter is to address these issues of contamination and detection of low amounts of protein in order to use GFP purified from the roots of *Arabidopsis thaliana* seedlings as a reporter protein for cell type specific metabolic activity.

The main objectives were to 1) develop a protocol that minimises biological and chemical contamination in the reporter protein samples, 2) identify the detection limits for ^{13}C labelling in small amounts of amino acids to achieve reliable detection, and 3) explore options for improving the detection of mass isotopomer distributions from samples containing low concentrations of amino acids.

4.1 Introduction

Measuring mass isotopomer distributions from amino acids derived from protein forms an important part of standard metabolic flux analysis experiments. It can be routinely performed from total protein extracts derived from relatively small amounts of biological material such as micro-organisms and plant cell cultures or tissue (Zamboni et al., 2009). Its reliability has been demonstrated by a variety of metabolic flux analysis experiments that have also discovered important features of plant metabolism through use of stable isotope labelling experiments and experiments based on simulated datasets (Allen et al., 2009a; Dauner and Sauer, 2000). However, GC-MS and all other mass spectrometric techniques have detection limits for small amounts of analytes. In the case of the reporter protein strategy, these limits are tested by the small amounts of specific protein that can be purified from the sub-sets of cells expressing the reporter protein.

The two published examples of validation of GFP as a reporter protein in micro-organisms both report a requirement for at least 1 mg of purified GFP in order to achieve reliable quantification of mass isotopomer distributions from the reporter protein (Rühl et al., 2011; Shaikh et al., 2008). Additional restrictions were described by Rühl and colleagues for the culture volume or ratio of the cell type of interest to other cells in the culture: less accurate flux ratios were obtained from two genetically modified strains with deletions that were shown to cause differences in central carbon metabolism, when the cell type of interest formed a smaller fraction of the total cell mixture. Allen and coworkers reported similar difficulties in reliable detection of some of the amino acids from purified plant RuBisCo (Allen et al., 2012).

The difficulties in measuring mass isotopomer distributions from small amounts of purified reporter proteins have three main components that are linked but can be addressed separately: contamination from biological background, contamination from sample preparation and the detection limits of the analytical techniques. Biological contamination in the samples can be either proteins from the total protein fraction in the purified GFP sample or other contaminants that are not fully removed through the purification steps. The first of these arises from insufficient washing steps or non-specific binding during purification and gives rise to the reported requirements to have a certain fraction of the cells

in the mixture expressing the reporter protein. Previous authors have attributed this to unspecific binding to the affinity matrix used for purification of GFP.

If the mass isotopomer distributions in two fractions of cells are different, contamination from proteins synthesised in the other cell type can distort the detected ^{13}C enrichment. The extent of this distortion will depend on the differences in the flux distributions in the two cell types and the relative contribution of the two cell types to the measured component. The first of these is unknown (since this is what is being investigated), while considerations regarding ratios of cells that must be expressing the reporter protein are inconvenient for analysis in multicellular organisms, where the numbers of cells of each type are set. Therefore the purification method used must be highly specific for the reporter protein in order to reduce the level of contamination from the total protein fraction to a negligible level.

The second type of biological contamination is other small molecules that are not sufficiently removed by the protein purification steps. In their original validation of the GFP reporter protein strategy in micro-organisms, Shaikh and co-workers describe the interfering effect of phosphate from the growth medium on the spectra measured from purified GFP (Shaikh et al., 2008). Molecules such as phosphate become problematic because they can be present at high levels in the original extracts, can be chemically derivatised and interfere with the detection of amino acid peaks.

Chemical contamination on the other hand can arise from the purification method itself. In the reporter protein work published to date, significant contamination has been reported by imidazole and urea introduced in the buffers used during purification (Shaikh et al., 2008), and amides released from polyacrylamide gels (Allen et al., 2012). Both necessitated changes to the original protocols. In order to achieve reliable detection of amino acid fragments from the reporter protein, the protocols used therefore have to be carefully considered and validated, as they may have a significant effect on the data obtained.

Steps taken to reduce biological and chemical contamination are a prerequisite for detection of reliable mass isotopomer distributions from amino acids derived from a reporter protein, but

ultimately detection limits of the analytical technique define the requirements for successful measurements. Other types of analytical instruments, beyond standard GC-MS, have been suggested as alternatives to enable more extensive analysis using reporter proteins. Both large volume injections (Rühl et al., 2011) and detectors with increased sensitivity (Allen et al., 2012) have been proposed. However, it is not clear to what extent either of these could provide solutions for overcoming the reported issues. Applying the reporter protein strategy to studies involving different cell types in plants is likely to increase the demands on sensitivity of the analytical equipment further.

In summary, published work has found the major challenges in applying a reporter protein strategy are (1) the small amount of protein that can be purified from sub-sets of cells, and (2) contamination from chemicals introduced during the purification steps or present in the biological background, which interfere with detection of the ^{13}C enrichment in amino acid fragments. The major aim of this chapter was to develop protocols for using the reporter protein strategy in plant cells taking into account these considerations. The specific objectives were to validate a highly specific system for purification of GFP and to develop a protocol for the reliable detection of the mass isotopomer distributions from small amounts of protein with GC-MS to allow investigation of cell type specific metabolism in the Arabidopsis root.

The purification of GFP from plant total protein extracts was achieved using a GFP nanotrap based on immunopurification using a nanobody (single domain antibody) derived from the sequence of a llama heavy chain antibody (Rothbauer et al., 2008). In addition, standard protocols were modified to allow detection of mass isotopomer distributions for amino acids from lower amounts of protein, and fragments that could be reliably measured from constitutively expressed plant GFP were identified. This chapter describes the different approaches that have been explored during the development of a protocol for reliable detection of small amounts of stable isotope labelled amino acids reliably and defines some of the analytical requirements for conducting analysis using limited amounts of protein. The work forms the basis for validation of the reporter protein strategy for investigating cell type specific metabolism in the Arabidopsis root.

4.1.1 Addressing biological and chemical contamination in samples with highly specific purification of the reporter protein

Purification of the reporter protein using a highly specific method is important to avoid contamination from other proteins potentially derived from other cell types in the tissue. In bacteria, the purification of GFP was based on a His-tag (Shaikh et al., 2008) or a glutathione s-transferase (GST) tag (Rühl et al., 2011). In both cases the protein was therefore purified with affinity purification: the tag consisting of several histidine residues was purified with nickel charged resin and the GST tag with a glutathione affinity resin. The introduction of a tag is an advantage that can only be achieved by using recombinant proteins and standard His-tag purification may still suffer from contamination by histidine rich proteins (Andersen et al., 2013). Nonspecific binding seemed to affect the purification with the GST tag when GFP was used to report on flux ratios in deletion mutants of *Escherichia coli* (Rühl et al., 2011). There are possible solutions for these problems, for example further size exclusion chromatography as suggested by the authors. However, the inevitable losses that occur during each purification step make this increasingly challenging as the amounts of protein are decreased.

The more specific option for purification is immunopurification with an antibody that binds specifically and with high avidity to the reporter protein. Purification with an antibody can be used for both native and foreign proteins with the downside of possible elution of the antibody heavy and light chains into the sample. This is problematic for metabolic flux analysis. It can be avoided by running protein samples on gel and further isolating the proteins of interest from antibody fragments. Previous work on RuBisCo exploited this method (Allen et al., 2012) and suggests reliable detection of mass isotopomer distributions from amino acid fragments derived from amounts of protein smaller than those required in the earlier studies on micro-organisms. Immunopurification therefore seems to have distinct advantages to affinity purification with a tag. On the other hand, this work also identified that contamination from the polyacrylamide gel affects the resulting spectra necessitating the introduction of further purification steps.

To achieve a high specificity purification without contamination from eluting antibody chains, the immunopurification of GFP was done using a GFP nanotrap (Rothbauer et al., 2008), where single chain antibodies are covalently bound to N-hydroxysuccinimide-activated Sepharose resin. The advantages of using this system include the highly specific purification of GFP using a single polypeptide, which can be efficiently expressed in *Escherichia coli*, facile purification based on elution of GFP off the resin with a change in pH, and the lack of co-eluting polypeptides in the sample. With the use of citrate buffer for elution, an additional advantage is avoiding imidazole, urea and glycine, which are present in some of the common buffers used in protein purification.

The variable single domain antibody fragments based on heavy chain antibodies from the Camelidae use a single heavy chain domain to bind to antigens, instead of forming antigen-binding sites between heavy and light chains. The GFP antibody is therefore small in size (approximately 13 kDa) compared to conventional antibodies and therefore often referred to as a nanobody (Rothbauer et al., 2008). The GFP binding protein (GBP) binds to GFP and YFP, but not to CFP or DsRed variants. This difference in binding is reportedly caused by the presence of an isoleucine residue at position 147 (Rothbauer et al., 2008). Nanobodies are inherently more stable than antibodies consisting of heavy and light chains, and may be further stabilised by introduction of disulphide bridges into their structure (Saerens et al., 2008). A specific loop protects the area that is involved in hydrophobic interactions with the light chain in the heavy chains of regular antibodies. These and other modifications increase stability of the structure such that the nanobodies can withstand high temperatures and a wide range of pH (with quantitative release of the antigen at pH 3.2), which makes them suitable for the often demanding conditions of protein purification (Rothbauer et al., 2008).

Potentially purification with the GFP nanotrap avoids several of the challenges associated with the effective isolation of reporter proteins. Immunopurification is highly specific to the antigen enabling extensive washing of the resin to allow effective minimisation of contamination by the total protein fraction. The co-elution of antibody heavy and light chains is avoided due to the single chain structure of the GFP binding protein. Finally, elution off the column can be achieved with a change

in pH, avoiding introduction of chemical contaminants known to cause problems with GC-MS analysis.

4.1.2 Amount of reporter protein required for chromatographic detection of ^{13}C mass isotopomer distributions from amino acids

The requirement to purify up to 1 mg of protein in order to measure reliable mass isotopomer distributions would set limits on the work that can be done in plant systems. However, the work on micro-organisms has not provided a very thorough exploration of the options that exist for detection of amino acid fragments from smaller amounts of protein. In plants, there is one example of published work where RuBisCO purified from *Brassica napus* embryos was used as a reporter protein and the stable isotope labelling was detected with GC-MS (Allen et al., 2012). This approach relied on an abundant protein that makes up large proportions of the total protein fraction. Although quantities of protein used for the analysis were not reported explicitly, information provided by authors on protein concentrations of the plant extracts suggests the absolute amounts of protein used were probably at least an order of magnitude lower than the detection limits reported for bacteria.

In the experiment with RuBisCo, 10 amino acids were detected in sufficient quantities to reliably measure mass isotopomer distributions with single ion monitoring (SIM). The detection limit described for GFP in work with micro-organisms might therefore be flexible, although Rühl and co-workers concluded through experiments that flux ratios obtained using data from 0.1 mg of GFP were inaccurate (Rühl et al., 2011). These differences might arise from RuBisCo being purified by immunopurification and the GFP by affinity purification using GST and from the use of single ion monitoring for detection of amino acids derived from RuBisCo.

If the reporter protein strategy is to be used on systems containing smaller proportions of GFP, such as specific cell types of the Arabidopsis root, the analytical challenge of detecting the largest possible dataset of amino acid fragments derived from low amounts of purified reporter protein remains and possibly amplified. Even with the careful purification of samples discussed earlier, contamination and molecules from reagents remaining after purification steps can affect the detection, if they are

present at considerably higher levels than the amino acids. The most abundant peaks in the analysis will limit how much sample can be injected without overloading the column and mass detector, and it may not be possible to remove all of these contaminating molecules from the samples. However, one option to combat this limitation may be to use selective methods such as selected ion monitoring (SIM), which measures the mass peaks for pre-defined masses only. This potentially increases the sensitivity of detection of those masses and reduces baseline noise. In SIM mode a mass spectrometer will use more time on detection of the selected masses, which can increase the sensitivity of the analysis by 10 to 100 times compared to scan mode by increasing the signal to noise ratio. However, this increase in sensitivity depends on time sharing and is achieved only by decreasing the number of fragments or ions detected in the analysis.

In order to obtain isotopomer data on an amino acid, several ions with different masses are measured simultaneously. Data are measured from several fragments containing different carbons, both for identification of the compound and for building a robust metabolic model. This will have a significant effect on the sensitivity obtained by running in SIM mode. To provide an example, Fialkov and colleagues describe testing the sensitivity increase achieved by SIM with a method using 3-ion SIM for detection of small amounts of analytes such as permethrin from a complex matrix (Fialkov et al., 2007). The first element that affects the sensitivity of SIM is time sharing, which for a three-ion method decreases the sensitivity by a factor of 1.73 (square root of three) compared to a one-ion method. However, due to effects of background noise on the least abundant fragments, the actual decrease in sensitivity was more likely to be closer to an order of magnitude. In order to achieve a high sensitivity measurement using a SIM method, it should therefore be run for a small number of ions at any specific time. In the instance of the reporter protein strategy this might be impractical for two reasons: 1) measuring isotopomer distributions requires data on a family of up to nine mass peaks relating to each amino acid fragment as well as at least one or two ions around them for confirmation of identity; and 2) these peaks arise at the same time and must be detected during the same run. Even if measurements of individual amino acids or their fragments are split across several runs to decrease the number of ions detected during one run, the sensitivity will not

be comparable to a one-ion SIM. The advantage over scan mode is then dependent on factors such as biological contaminants in the sample. In the case of permethrin a complex background decreased the limit of detection from 1 pg to 10 pg for a two ion SIM method (Fialkov et al., 2007).

The best option is to try to inject greater quantities of amino acids into the analysis. Ultimately the levels of amino acids detected depend on the concentration and volume of the sample injected. Higher quantities of identifiable fragments can be achieved either by increasing the concentration of the sample or by increasing the injection volume. If the amounts of amino acids are low in the sample, the volume of derivatisation reagents required may be considerably lower than for standard techniques. The benefit of concentrating samples this way is of course dependent on the level of contaminants in the sample: as the GC-MS analysis itself is not selective for amino acids (even in selective modes high levels of contaminants are not desirable as they will have to pass through the system), any contamination in the samples will be equally concentrated and may therefore still overload the system. Another limitation is that the sample volume has to be large enough to allow injections.

The second option is using larger injection volumes. In regular inlets, the limit for the maximum injection volume is defined by the volume of the solvent used in samples in gas form, which must fit within the liner. Therefore large volume injections normally require a specific type of inlet, where the solvent can be purged by defining when the first analyte of interest is eluted. This effectively concentrates the sample after injection, as all compounds prior to the point of elution of the first ion of interest are removed from analysis. The advantages of this are that it does not affect the sample preparation and can reduce contamination in samples. Even if 1 mg of protein is purified from samples, the final steps of sample preparation dilute the sample so that with 1 µl injections only approximately 14 µg (hydrolysed into amino acids) enters into the analysis. Most large volume injection inlets allow injection of a range of volumes that can exceed the total volume of normal derivatised samples.

Finally, the sensitivity of the analytical instrument sets the ultimate limits of detection and guides the development of other methods and protocols around the detection. In order to measure mass isotopomer distributions reliably, the amount of contaminants from sample preparation and biological background should not be so high they will saturate the detector and the levels of amino acid fragments detected should be high enough for all the necessary mass peaks to be distinguishable from baseline noise. Several of the analytical aspects discussed in this introduction were tested in order to develop a protocol for reliable detection of amino acids derived from GFP expressed in the Arabidopsis roots.

4.1.3 Reliable detection of mass isotopomer distributions from small amounts of amino acids with GC-MS

There is a difference between detecting an amino acid based on the most common ion fragments that are produced in ionisation and quantification of mass isotopomer distributions. The difference arises from the fact that ^{13}C enrichment is detected from mass peaks, which have respectively higher masses based on how many atoms of the heavier isotopes they contain. For example, the M+1 mass peak measures the total ion count of fragments that contain one ^{13}C substitution, whereas M+2 measures fragments that contain two. The number of peaks to be detected depends on the size of the carbon skeleton of the specific fragment.

In samples with natural abundance of ^{13}C the number of molecules that would contain a large number of ^{13}C substitutions is close to zero. Here variation can arise if the levels of the M+0 peak are already low and baseline noise gets interpreted as ^{13}C enrichment in the heavier mass peaks or the increase is lost in the baseline. The opposite is true if the substrate is labelled with high levels of ^{13}C, so it is more likely to generate molecules containing more ^{13}C than ^{12}C. In more even mixtures, for example in cultures with 20% [$^{13}C_6$]-glucose, the distribution of ^{13}C is slightly more even. Here a larger number of peaks are significantly different from zero, but may still have quite low total ion counts. For positionally labelled molecules such as 99% [1-^{13}C]-glucose, variation is again different, depending on which carbons from the substrate contribute to the detected fragment and which

pathways have been used to synthesize it. Therefore the ability to detect mass isotopomer distributions depends on the ability to accurately distinguish each of the mass peaks from the baseline noise.

Other factors that give rise to unreliable mass isotopomer distributions are overlap with contaminants and low levels of a specific fragment. It is not the level of the amino acid peak observed in the total ion chromatogram that determines how reliable the detection of mass isotopomer distributions for the fragment is, but the levels of the mass peaks within a fragment. The latter needs to reach high enough levels for each of the fragments used in analysis. The reliable detection of mass isotopomer distributions from GFP in the Arabidopsis root therefore relies on first detecting high levels of amino acids from the samples and then detection of reliable mass isotopomer distributions at high enough levels from the fragments derived from those amino acids. The following sections describe work on identifying what these levels are in practice and which fragments can be reliably detected from the samples obtained from GFP from the Arabidopsis roots.

4.2 Results

4.2.1 Highly specific purification of GFP using a GFP nanotrap

4.2.1.1 Expression and purification of His-tagged GFP binding protein

The GFP binding protein derived from the llama single heavy chain antibody fragments can be expressed in *Escherichia coli* using the expression plasmid pHEN6 (Rothbauer et al., 2008). After initial expression in the WK6 strain, the BL21 strain was found to give a higher yield of functional protein. Suitable expression conditions were found during previous work with the nanobody by testing expression in different media at different temperatures. The conditions used in this thesis are described in more detail in Chapter 2.

The size of the GFP binding protein (GBP) is approximately 13 kDa (Rothbauer et al., 2008). The protein is expressed in the periplasm and in order to obtain the periplasmic fraction, osmotic shock was carried out prior to His-tag purification with nickel charged resin (Figure 9). The combined extraction from the periplasm and purification with metal affinity chromatography resulted in yields of approximately 2 mg of protein from 500 ml of culture, measured by Bradford assay using BSA as

a standard (Figure 10). Subsequent purification by either column chromatography or batch-elution from bulk resin was effective. Approximately 1 mg of purified nanobody protein is required to bind to 1 ml of the NHS-activated Sepharose to produce the antibody resin. Therefore approximately 4 ml of resin could be produced from GBP protein expressed in 1 l of culture medium (sufficient for 80 GFP purifications).

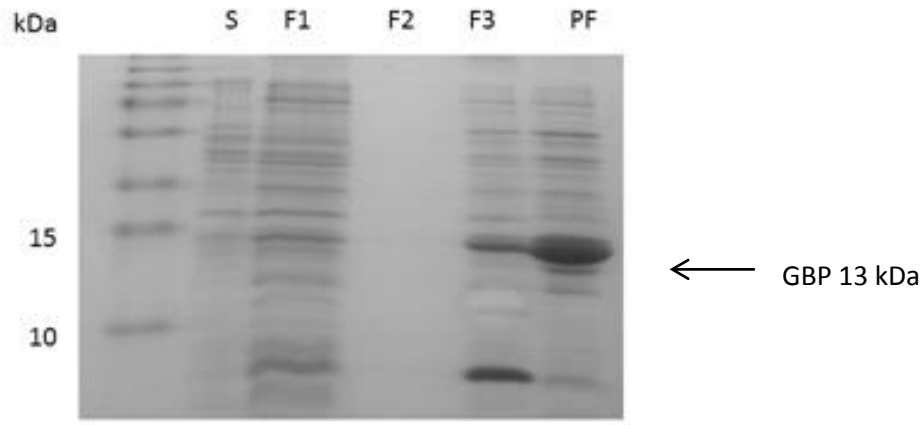


Figure 9: **GFP binding protein expression in BL21 cells.** Samples corresponding to the different steps of treatment obtained from osmotic shock of BL21 cells containing the pHEN6-GBP plasmid were analysed by SDS-PAGE. Cells were grown in 500 ml of Terrific Broth with 0.2% glucose and induced with IPTG. SDS-PAGE analysis of the different fractions was run on 12% gel. The fractions were: S, supernatant from first centrifugation step; F1-3, fractions 1-3 (supernatant); PF, periplasmic fraction. The periplasmic fraction was concentrated using a spin column and should contain the llama single chain antibody that is approximately 13 kDA in size. A band of this size is indicated with an arrow.

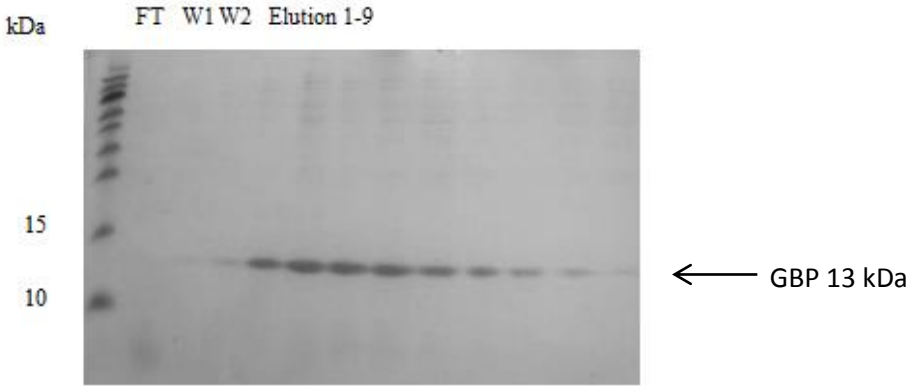


Figure 10: **Purification of the His-tagged GFP binding protein from the periplasmic fraction.** Samples obtained during metal-affinity chromatrography of the periplasmic fraction of containing GFP were

analysed by SDS-PAGE. Protein was bound from the periplasmic fraction to Ni-NTA resin, which was washed twice to remove remaining supernatant. The protein was eluted using consecutive 3 ml volumes of imidazole buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole pH 8.0). Samples from the different eluate fractions were concentrated with a spin column and the resulting samples were run on an SDS-PAGE gel to confirm the presence of the single chain antibody with the approximate size of 13 kDa, indicated by an arrow. FT, flow through from the resin; W1-2, washes.

4.2.1.2 Purification of GFP from plant total protein extracts

The purification of GFP using the antibody bound to the NHS-activated resin was tested with extracts obtained from roots of transgenic Arabidopsis seedlings that constitutively expressed GFP. Elution buffer was changed from glycine typically employed for release of antigens from antibody-linked resins during immunoaffintiy purification to citrate, to avoid contamination by unlabelled amino acids. To test the specificity of GFP binding, a protein sample from the same amount of tissue from Columbia wildtype seedlings was subjected to the same procedure. Since these seedlings did not contain GFP, any protein from these latter samples specifically eluting from the antibody column provides an indication of likely contamination from other proteins in the initial plant extracts.

The protein sample specifically eluting from the immunoaffinity column contained a band of a size corresponding to that of GFP (approximately 27 kDa) when analysed by SDS-PAGE (Figure 11). In addition, no bands were visible in the sample purified from the root extracts from plants that did not contain GFP. The band was faint, suggesting the levels of protein purified were near the lower limits required for SDS-PAGE analysis: typically visualisation on Coomassie Blue stained gel requires 30-100 ng of protein (Simpson, 2010). However, only a part of the eluate was run on gel as the lowest sample volume achieved with a spin column was approximately 20 µl and with added dye, it was not possible to load all of this on the gel. Two washes were not necessarily enough to remove all the traces of the total protein fraction, but in five washes there were no remaining traces of protein contamination either in the wash fraction or in the Columbia total protein sample.

No amino acid fragments were detectable in GC-MS analysis of a reagent blank (Figure 12) or in samples purified from wildtype Columbia seedlings that did not express GFP (Figure 13). The lack

of amino acid peaks in the specific eluate obtained from the wildtype seedlings indicates that the purified GFP obtained from seedlings constitutively expressing this protein were unlikely to be appreciably contaminated by other proteins. This result agrees with analysis on SDS-PAGE gels. Amino acids were, however, detected from the GFP sample (Figure 14). Equivalent results were obtained when the experiment was later repeated with protocols that allowed detection of amino acids derived from GFP at higher levels. Together these results show the GFP nanotrap purification is highly specific for GFP and can be used to purify GFP from plant protein extracts.

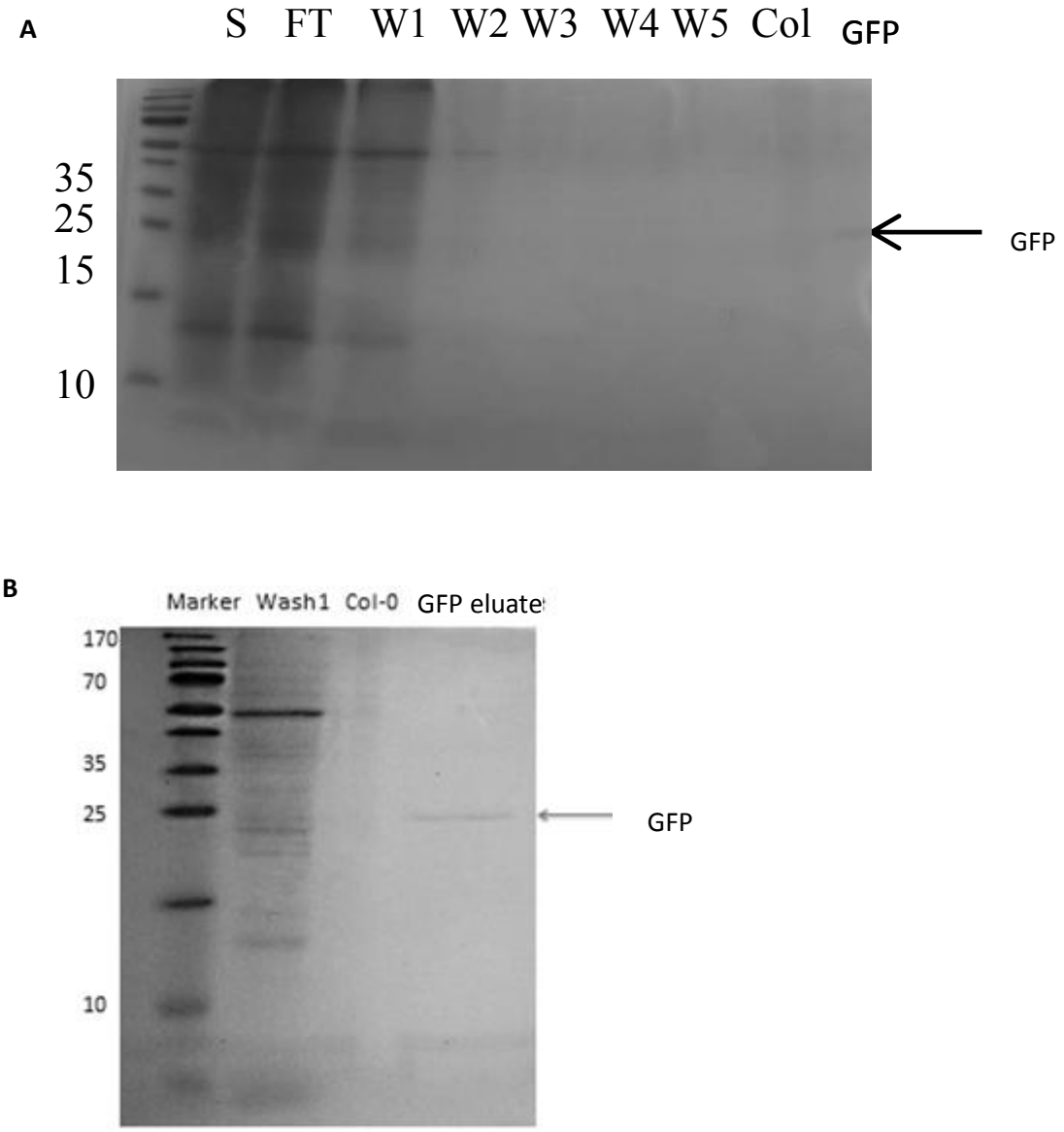


Figure 11: **Purification of GFP from plant total protein extracts using the GFP nanotrap.** GFP was purified from total protein extract from transgenic plants with constitutive GFP expression and wildtype plants that do not express GFP. The antibody resin with GBP bound to NHS-sepharose was incubated with the total protein extracts, which were subsequently washed five times to remove all contamination from the total protein fraction. GFP was eluted with citrate buffer and concentrated with a spin column. The resulting samples were run on 12% SDS-PAGE gel together with fractions from all the purification steps. Samples are: S, Original extract (supernatant); FT, flow thorough; W1-W5, washes; Col, eluate from Columbia wildtype; and GFP, eluate from a line with constitutive GFP expression. Two separate purifications are shown in panels (A) and (B) and the faint GFP bands are indicated with an arrow. Panel (B) shows a band that was identified as GFP with peptide mass fingerprinting (Apendix 3).

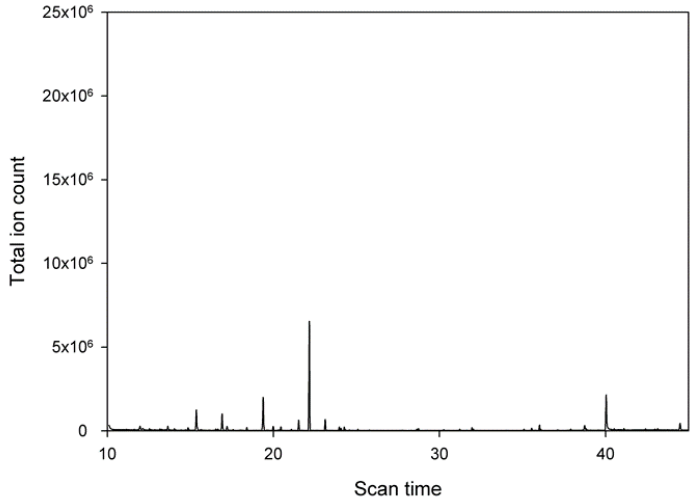


Figure 12: **Total ion chromatogram of the derivatisation reagents used in GC-MS analysis of protein hydrolysates.** A reagent blank lacking added amino acids was treated with tBDMS in pyridine using the standard amounts and conditions employed in conventional analysis of amino acids.

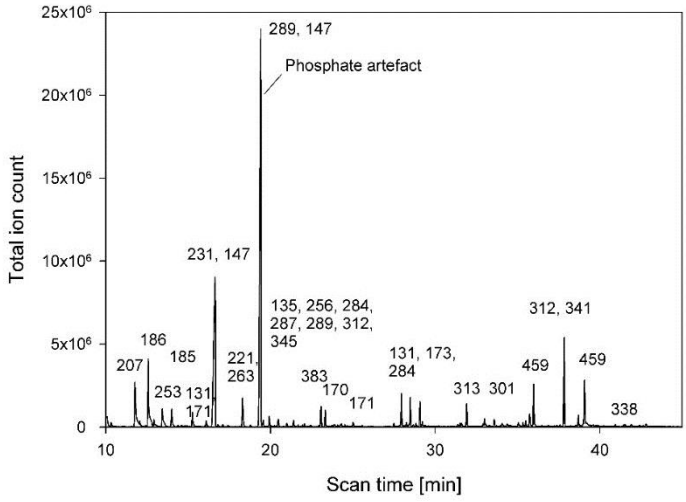


Figure 13: **Chromatographic analysis of a sample purified from wildtype seedlings using the GFP nanotrap**. Protein was extracted from one culture of 30-50 Columbia wildtype seedlings that do not express GFP, immunopurified with the GFP nanotrap, and any protein retained by the immunoaffinity resin hydrolysed and derivatised. The standard conditions were used, except for the total volume of the derivatisation reagents, which was reduced to one third of normal to allow detection of trace levels of amino acids during subsequent GC-MS analysis Mass/charge ratios of the ions that form the main peaks are shown and do not correspond to any fragments characteristic of amino acids.

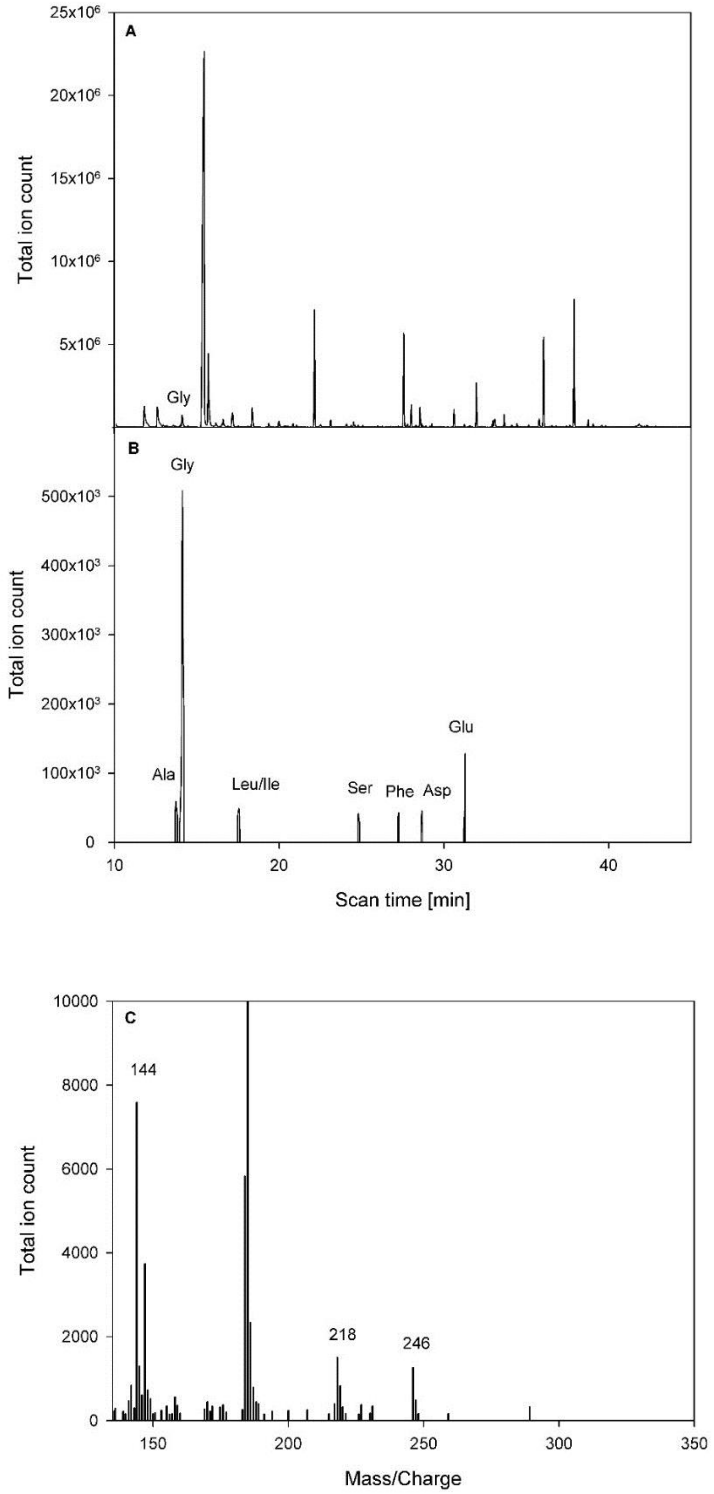


Figure 14: **Standard chromatographic analysis of a hydrolysed GFP sample.** GFP was immunoaffinity purified from seedlings constitutively expressing the reporter protein, hydrolysed and then derivatised using the normal amount of tBDMS in pyridine under conventional conditions. (A) Total ion chromatogram of GFP immunopurified from a single culture of 30-50 seedlings. Glycine elutes in the peak labelled Gly. (B) Edited total ion chromatogram of immunopurified GFP in which spectral regions not containing fragments derived from amino acids have been removed, allowing scale expansion to facilitate

visualisation of the amino acid peaks. (C) Mass spectrum of glycine containing peak in which diagnostic fragments (m/z: 144, 218 and 246) are indicated against large quantities of contaminating ions. Mass spectra of the other amino acids detected in the sample are presented in Error! Reference source not found..

4.2.1.3 Protein identification with peptide mass fingerprinting

Peptide mass fingerprinting is a widely used tool for identification of proteins through the detection of peptide sequences that are present after a purified protein is digested by a proteolytic enzyme. The resulting mixture of peptides is analysed by mass spectrometry to determine their molecular weights. This size of these peptide fragments is then compared against a database to identify proteins that produce matching weights when subjected to a digest with the same protease.

A GFP sample was immunopurified from Arabidopsis seedlings with constitutive GFP expression and run on SDS-PAGE gel. The major band present on the gel was subjected to trypsin digestion in situ and the resulting peptide mixture was analysed by the Central Proteomics Service at the University of Oxford by time-of-flight mass spectrometry. The MS/MS ion search was done using standard parameters with monoisotopic mass values, unrestricted protein mass, peptide mass tolerance of ± 20 ppm, fragment mass tolerance of ± 0.5 Da and maximum missed cleavages limited to three. The database the measurement data was searched against was NCBInr and taxonomy *Arabidopsis thaliana*. The significance threshold was set as $p < 0.05$. To limit the matches the ions score cut-off value was set as 20. The protein was identified as GFP on the basis of 11 peptide matches that exceed the threshold values for 95% confidence. Protein sequence coverage for the matched peptides was 31% (further details of the analysis are presented in Appendix 4).

4.2.1.4 Quantification of GFP purified from plant seedlings using fluorescence

The fluorescence of GFP was used to quantify levels of the reporter protein in root and leaf tissue derived from plants with constitutive expression of GFP. The fluorescence standard used was a His-tagged GFP expressed in *Escherichia coli*, which was purified using metal-affinity chromatography on a Ni-NTA resin. The protein concentration in the purified GFP was quantified using the Bradford assay with BSA as a standard. The quantities of GFP used for calibration of the fluorescence

measurements ranged from nanograms, where the measurements were found to be less reliable, to microgram. The measured fluorescence (in arbitrary units) increased with the amount of standard, as expected (Figure 15). The amount of GFP purified from root total protein extracts of plants with constitutive expression of GFP was 700 ng (±570 ng) per culture of 30-50 seedlings. On average leaf tissue constituted 1.7 (± 0.91) mg and root tissue 0.7 (± 0.57) mg of the total weight of plant material in these cultures. The variation is in each case expressed as standard deviations. The levels measured for GFP purified from leaves were higher than for roots (1.67 μg ± 0.91 μg), but where levels of GFP purified from leaves were high, the levels purified from roots were similarly higher than in other samples (Figure 16). The total weight of tissue obtained from the liquid cultures varied between 0.6 and 5.1 mg. The variation in weight was mainly due to the differences in number of seedlings in culture after seed sterilisation and germination of the seeds.

To confirm the efficiency of binding with the custom-produced GFP nanotrap was comparable to commercially available nanobodies and to investigate the levels of GFP that could be purified from lines with cell types specific expression, a commercial microplate with GFP binding protein bound to the wells was used to quantify GFP purified from the roots of the Arabidopsis line with constitutive expression and four lines with cell type specific expression of GFP. Roots from five cultures of 30-50 seedlings were extracted and GFP from the combined samples was bound to the nanobody in the wells. The fluorescence measurements again showed the level of GFP purified from the line with constitutive expression was on average 749 ng (±274 ng) per culture (Figure 17). For the other lines, the quantities of GFP varied from less than 10 ng per culture to 530 ng per culture. This variation can in part be explained by different numbers of cells expressing the reporter protein. Based on the purification from the line with constitutive expression, the commercial nanobody and custom-synthesised antibody resin bound equivalent amounts of GFP from root extracts.

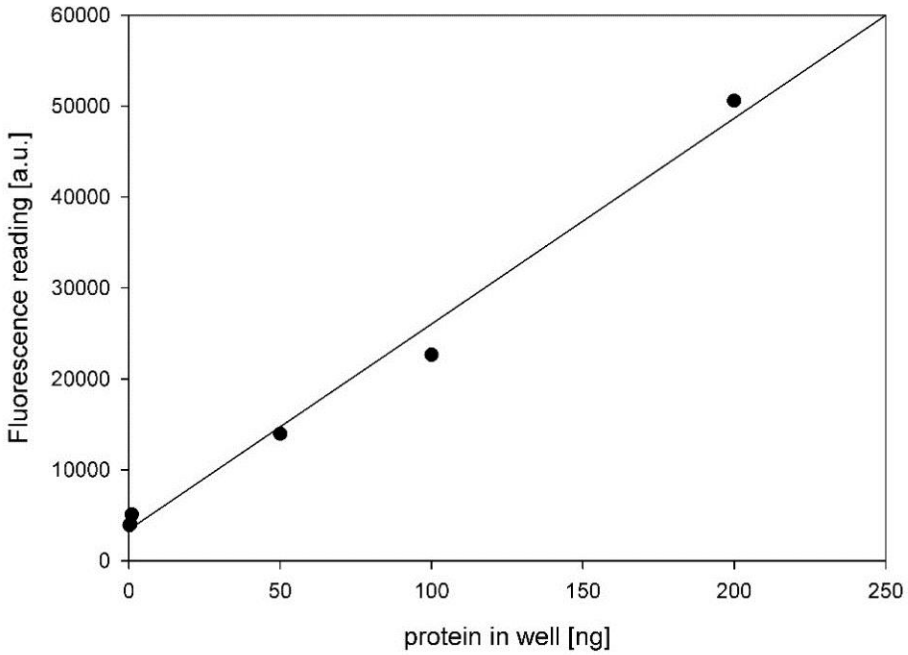


Figure 15: **Fluorescence measurements using GFP standard expressed in *Escherichia coli*.** Fluorescence was measured (in arbitrary units) using a range of concentrations of purified GFP. The protein content of the GFP standard was determined with the Bradford assay using BSA as a standard. Measurements were taken from a 96-well microplate in a total volume of 100 µl. Shown here are measurements from standards up to 200 nanograms. Measurements over a broader (microgram) range are presented in Appendix 5. Each plotted value is the mean of three measurements. Standard deviations are covered by the symbols.

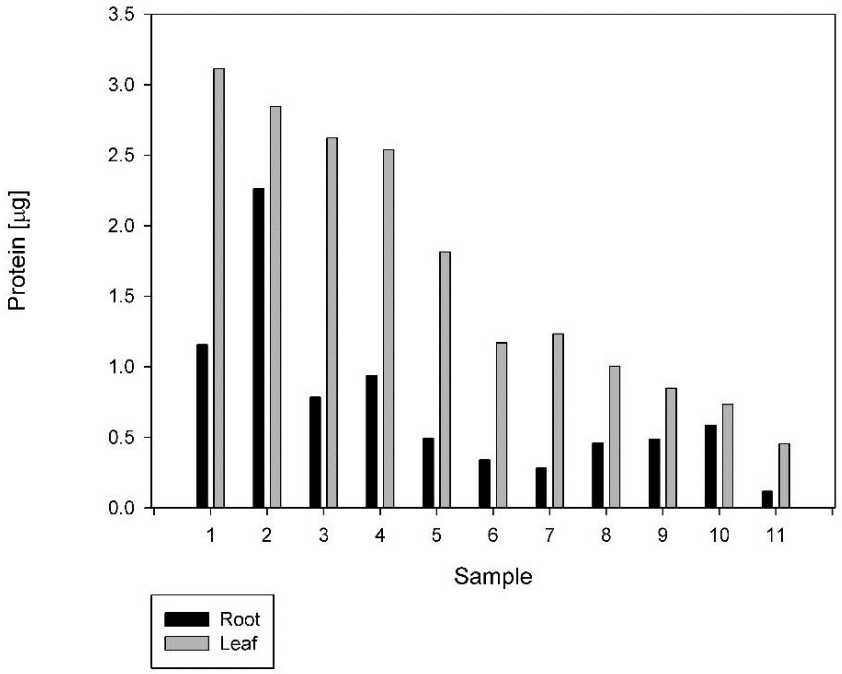


Figure 16: **Quantification of GFP purified from leaf and root tissue of Arabidopsis lines with constitutive expression.** Leaf and root tissue was separated from Arabidopsis cultures of 30-50 seedlings with constitutive GFP expression and GFP was purified using the GFP nanotrap. Fluorescence was measured on a 96-well microplate in a total volume of 100 µl. Each value is the mean of three separate measurements. The protein concentration was determined by comparison to a calibration curve constructed using known amounts of a purified GFP standard.

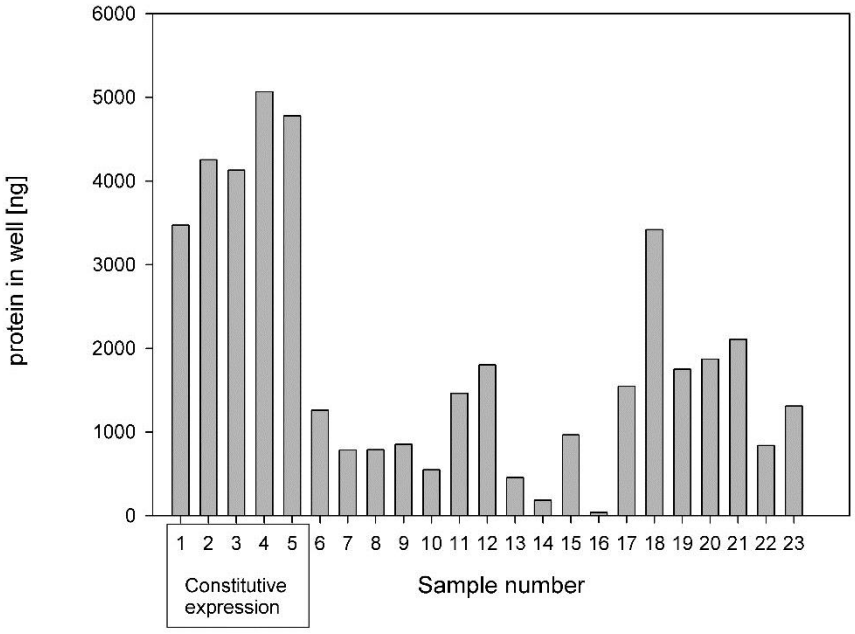


Figure 17: **Quantification of GFP purified from lines with constitutive or cell type specific patterns of expression on a commercial 96-well microplate coated with GFP binding antibody.** GFP was purified from five cultures of 30-50 seedlings of Arabidopsis with constitutive expression of GFP (samples 1-5) or cell type specific expression of GFP (samples 6-23). The lines with cell type specific expression had expression in stele, endodermis, cortex or epidermis cells. Fluorescence measurements were carried out on a commercial 96-well microplate coated with GFP binding antibody.

4.2.2 Reliable isotopic analysis of amino acids derived from GFP immunopurified from Arabidopsis seedlings

4.2.2.1 Detection of amino acids of GFP purified from plant seedlings with GC-MS

Amino acid standards of known concentration run with Agilent 7890A GC-MS instrument can be used to estimate the minimum amount of GFP required for detection of each amino acid (Appendix 6). Approximately 20 ng/µl concentrations of amino acids were required for detection of the majority of the amino acids at levels above 100,000 total ion counts, whereas arginine, histidine, cysteine and tryptophan were not detected at that level even at near to 200 ng/µl concentrations. The amino acid quantities measured by GC-MS increased linearly according to the concentration (following normalisation against an internal standard, norvaline).

In order to detect amino acids from the GFP extracted from seedlings with constitutive expression, a sample of GFP purified from the seedling extract was hydrolysed, derivatised under conventional conditions and analysed with GC-MS. Results indicate that the quantities of GFP based on the levels of amino acids detected from the samples in GC-MS analysis agree with those measured by fluorescence (Appendix 5). Using the standard analytical protocols, amino acids such as glycine and isoleucine were detected near trace levels from the plant GFP, suggesting levels lower than 5 ng/µl. Based on the fluorescence measurements the amount of hydrolysed protein in this type of GFP sample would be approximately10 ng/µl, but levels of each amino acid depend further on derivatisation efficiency, stability during hydrolysis and the number of residues in the specific protein (amino acid composition of GFP is shown in Appendix 7).

More promising results were obtained when the amount of biological material used for each immunopurified sample was increased to a total of three seedling cultures of 30-50 seedlings. The mass spectra corresponding to the normal results from this type of analysis are presented in Appendix 8. While this is ultimately the best approach for achieving consistent detection, increased workload and costs limit the feasibility of the increase in the number of cultures that would be required to examine cell-type specific expression lines that contain considerably lower levels of the reporter protein.

4.2.2.4 Detection of mass spectra from amino acid fragments derived from GFP purified from plant seedlings

The two factors that affect the amount of analyte in the analysis are the concentration of the analyte in the sample and the volume of sample analysed (which is determined by the injection volume in the case of GC-MS). Since the samples of protein hydrolysate are dried before derivatisation, the concentration in the amino acid samples is determined by the volume of derivatisation reagents used. Preliminary analysis established that decreasing the volume of the final sample could be an effective way of increasing the levels of amino acids detected in a fixed volume of sample. The ratios of solvent and the tBDMS reagent were kept constant, while the total volume decreased.

The smallest volumes that can be used are defined by several factors. Any contamination in the samples that is at higher levels than the amino acids can lead to overloading of the column or detector. The volumes of reagents must also exceed those required for efficient derivatisation of the amino acids. Finally, the total volume should be high enough to take into account evaporation of sample at different steps and to make sure no air is injected into the system. Within these minimum criteria, a series of derivatisation volumes was compared for hydrolysed protein (Figure 18). The optimum volume that allowed comfortable sample preparation, but still yielded high levels of detected amino acids was approximately 20 µl. However, when the total volume was reduced, some of the highest peaks were reaching saturating levels, suggesting the volume should not be reduced much further (Appendix 9). Using this optimised protocol, amino acids could be detected at significantly higher levels (Figure 19 and Appendix 10 for mass spectra of each individual amino acid detected).

To bring concentrations of amino acids in samples to a level that could potentially give reliable mass isotopomer data, GFP purified from three cultures of 30-50 seedlings was combined in one sample and run using the optimised protocol (Figure 20). This resulted in detection at levels of over 100,000 total ion counts for the majority of the amino acids (Appendix 11). The fractional abundances of ^{13}C detected from the amino acid fragments were at approximately 20% and further analysis of the labelling and mass isotopomer distributions is presented in Chapter 5.

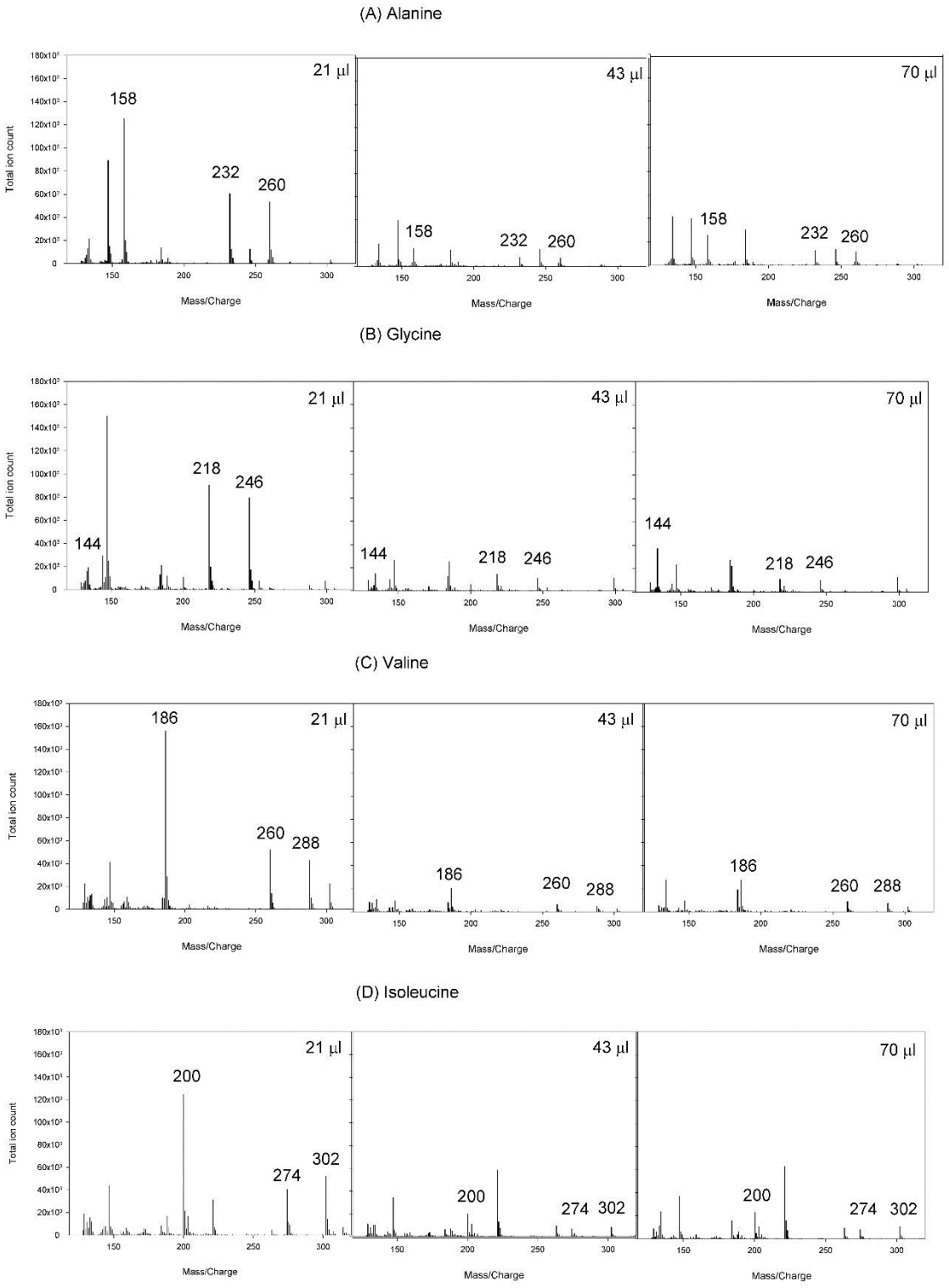
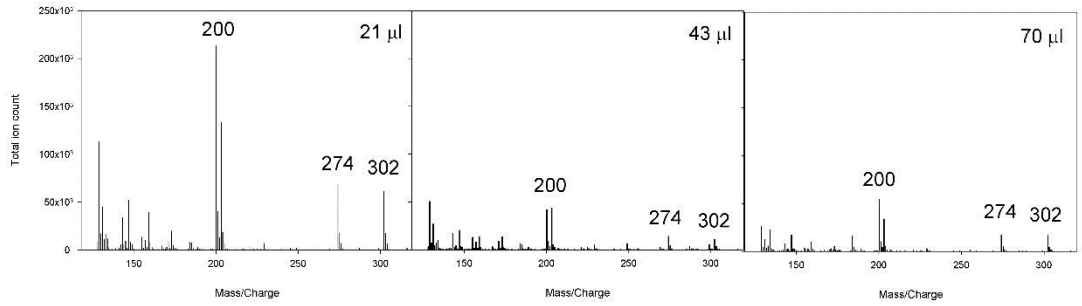


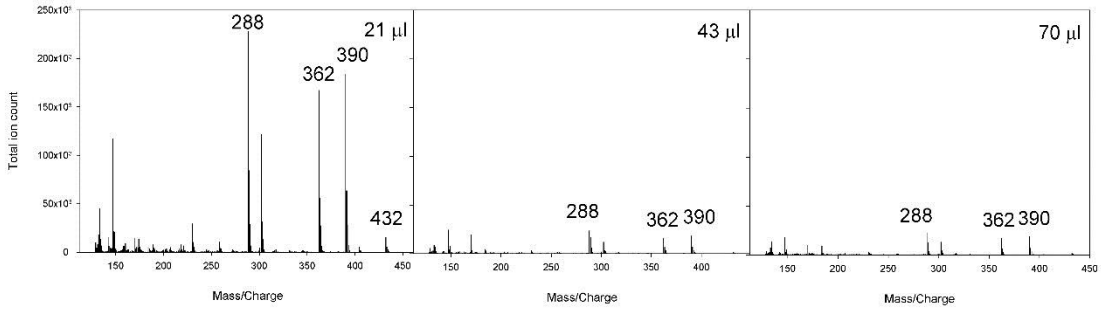
Figure 18: **Mass spectra of GFP samples immunopurified from 30-50 seedlings with constitutive GFP expression, hydrolysed and derivatised using varying quantities of derivatisation reagents.** Samples with total volume of 70 µl consisted of 30 µl of pyridine and 40 µl of tBDMS reagent. The ratios of the two reagents were kept the same across different volumes and were based on the standard protocol for derivatisation of amino acids from hydrolysed proteins samples.

Detection of arginine, histidine, cysteine, methionine and tryptophan was not consistent from these GFP samples and data is omitted from this comparison.

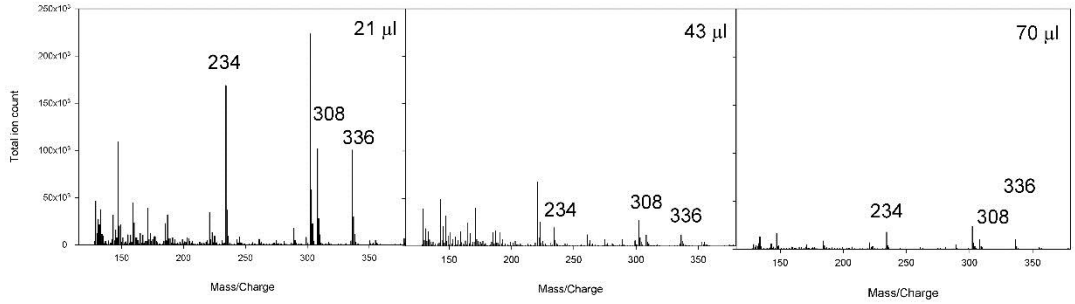
(E) Leucine



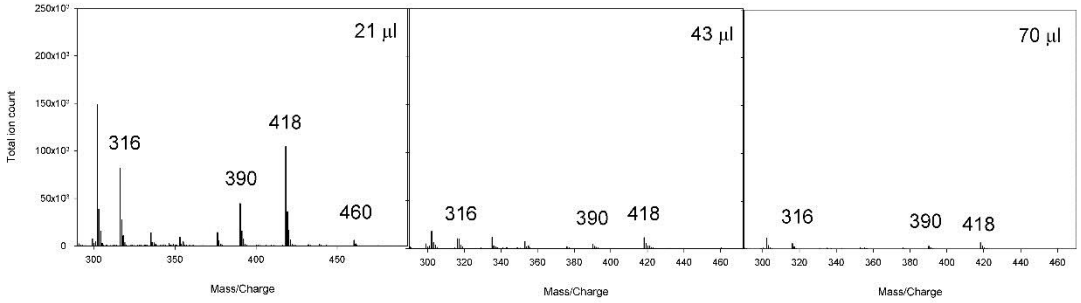
(F) Serine



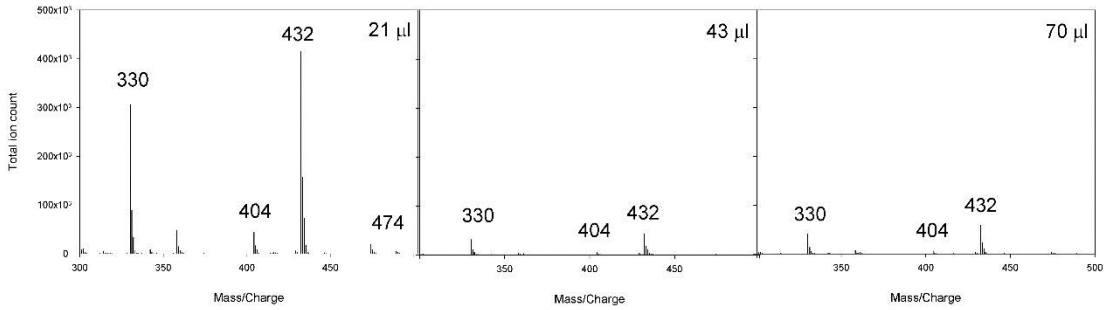
(G) Phenylalanine



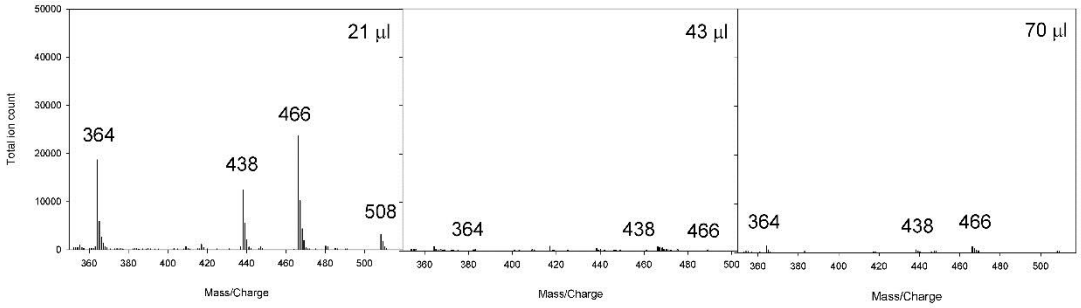
(H) Aspartate



(I) Glutamate



(J) Tyrosine



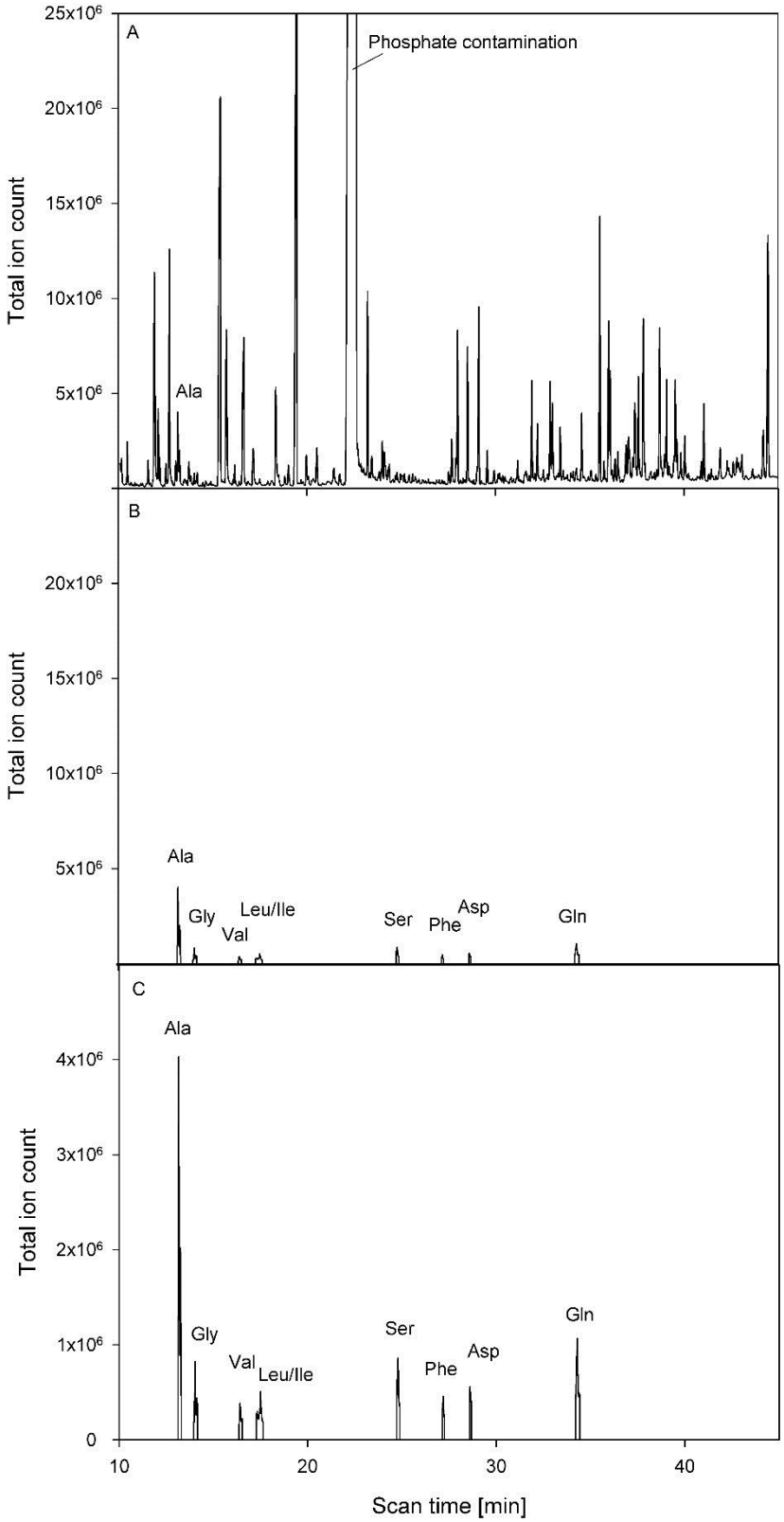


Figure 19: **Chromatographic analysis of a GFP sample immunopurified from a single culture of 30-50 seedlings with constitutive GFP expression, hydrolysed and derivatised using a reduced volume of reagents (20 µl).** GFP was immunoaffinity purified from seedlings constitutively

expressing the reporter protein, hydrolysed and then derivatised using reduced amounts of tBDMS in pyridine under conventional conditions. (A) Total ion chromatogram of GFP immunopurified from a single culture of 30-50 seedlings. Alanine elutes in the peak labelled Ala. (B) Edited total ion chromatogram of immunopurified GFP in which spectral regions not containing fragments derived from amino acids have been removed, to facilitate visualisation of the amino acid peaks. The scale has been kept the same as in A. (C) Edited total ion chromatogram of immunopurified GFP in which spectral regions not containing fragments derived from amino acids have been removed. Scale has been adjusted for visualisation of approximate total ion counts corresponding to peaks. Corresponding mass spectra are presented in Appendix 10.

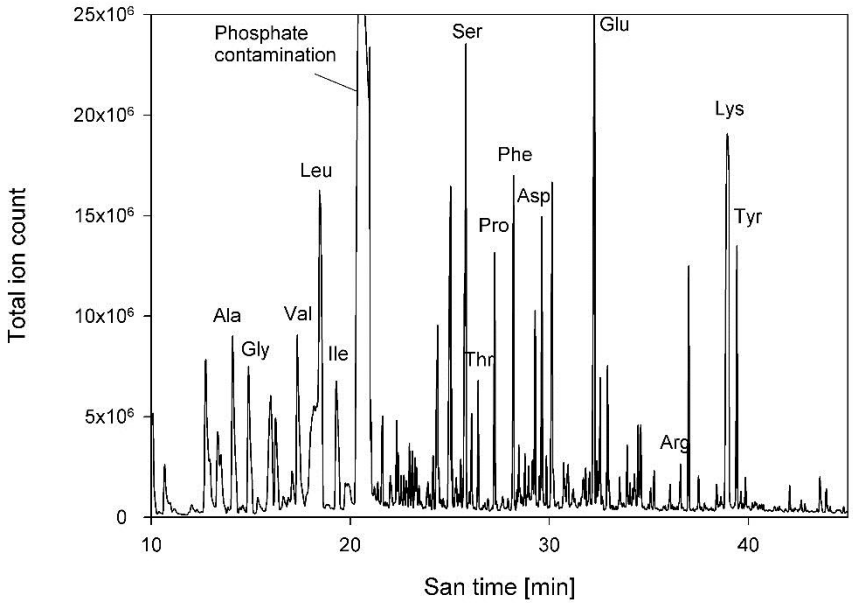


Figure 20: **Chromatographic analysis of GFP immunopurified from three cultures of 30-50 seedlings with reduced volumes of reagents (20 µl).** GFP was immunoaffinity purified from three cultures of 30-50 seedlings constitutively expressing the reporter protein, hydrolysed and then derivatised using reduced amounts of tBDMS in pyridine under conventional conditions. Total ion chromatogram with peaks containing amino acid fragments labelled with the corresponding three letter code is shown. Corresponding mass spectra are presented in Appendix 11.

4.2.1.3 Addressing phosphate and citrate contamination in immunopurified samples

Contamination is present in biological samples to varying extents. In the purified GFP samples, the main components present at high enough levels to affect the analysis were phosphate derived compounds and citrate introduced in the elution buffer. Concentrating the samples increased the potential for interference by these two main contaminants (Appendix 9). The first of two purification steps to address this contamination was to desalt the protein sample by ultrafiltration using an Amicon Ultra centrifugal filter. This step reduced citrate contamination, but was insufficient to remove the phosphate derived contaminants.

Secondly, in order to remove the remainder of the potentially interfering contaminants, the GFP samples were further purified by solid phase extraction filtration using a reverse phase C_{18} ziptip. These tips contain a 0.6 µl bed of C_{18} material fitted on to a pipette tip and are routinely used for sample purification in a range of mass spectrometry applications. By binding only the protein in the samples and using very pure reagents, the only remaining contaminants after filtration were those that could also be detected in a blank consisting of derivatisation reagents. The added filtration step was tested for loss of protein by filtering an aliquot of a protein standard sample (1 µg/µl final concentration, BSA) and comparing the GC-MS spectra of amino acids after a protein hydrolysis step to that obtained from an identical aliquot that was not filtered prior to hydrolysis and analysis (Figure 21). There was significant loss due to filtering: the total ion counts indicate the loss of amino acids was in the range of 30%. The biological samples purified with solid phase extraction had negligible levels of chemical or biological contaminants, but still contained relatively high levels of amino acids (Figure 22). The only major limitation of the method is that each tip can bind a maximum of 5 µg of protein thereby limiting the potential yield of a single tip. This should be taken into account if the combined quantity of purified GFP from the number of cultures used is likely to be near or over the limit, in which case multiple tips may be used.

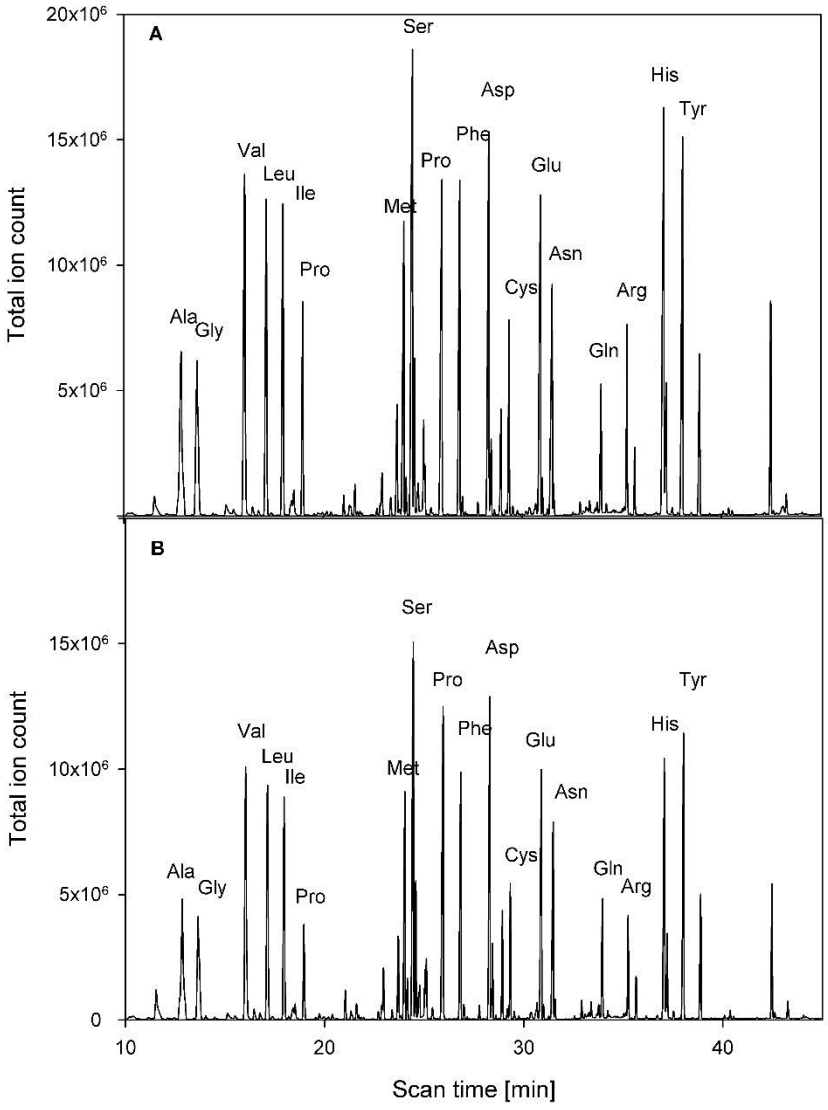


Figure 21: **Chromatographic analysis of the recovery of protein following purification by solid phase extraction.** Half of the total volume of a BSA standard (1 µg/µl final concentartion) was purified with a C_{18}-filter. Both halves of the sample were hydrolysed, derivatised and analysed with GC-MS using conventional conditions. (A) Total ion chromatogram of a hydrolysed and derivatised BSA standard. (B) Same BSA standard after solid phase extraction with C_{18}-filter. Sample was hydrolysed and derivatised after purification.

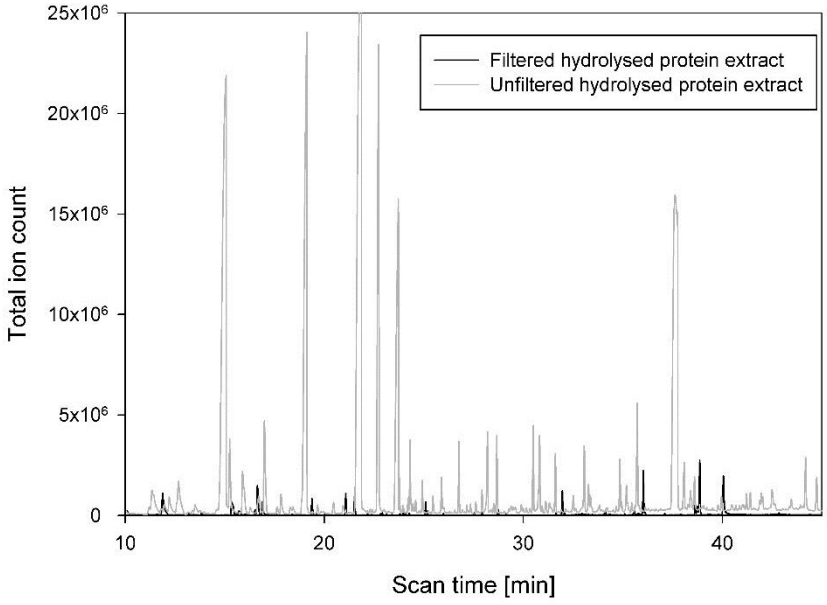


Figure 22: **Chromatographic analysis of protein extract from Arabidopsis seedlings before and after solid phase extraction.** Total protein was extracted from the roots of Arabidopsis seedlings. One sample consisting of a 200 µl fraction of the whole sample (2 ml) was purified with solid phase extraction. Both samples were hydrolysed and derivatised using conventional conditions. The overlay shows comparison of the total ion chromatograms from the two samples.

4.2.3 Detection of mass isotopomer distributions from the reporter protein

4.2.3.1 Reliable detection of mass isotopomer distributions from amino acids derived from protein hydrolysates

Determination of mass isotopomer distributions from mass peaks that have been detected at levels close to the limit of reliable quantification gives rise to significant variation. This is most likely due to the effect of baseline noise on quantification of the low mass peaks (Wittmann, 2007). In order to identify the minimum signal levels required for reliable detection, the detected total ion counts for 11,500 amino acid measurements from plant total protein and GFP were plotted against the calculated fractional abundance of ^{13}C. The data consisted of samples from plants with natural abundance of ^{13}C or grown in either 20% [$^{13}C_6$]-glucose or 80% [$^{13}C_6$]-glucose. The detected levels

of amino acids varied widely between samples, but notably contained a large number of measurements containing low total ion counts.

To identify any general patterns that apply across amino acids and fragments, the abundances or total ion counts for each fragment were compared with the fractional abundance of ^{13}C that had been calculated from the measurement data (Figure 23).The comparison showed random levels of ^{13}C are detected at low fragment abundances, but that the expected approximate levels of ^{13}C enrichment (natural abundance, 20% and 80% provided in substrate) start to emerge at total ion counts above 50,000. However, consistent separation between the three sets of values is only achieved when the abundance of the fragment exceeds several hundred thousand counts and even at the highest levels there are a few consistent outliers. (However, since the latter data points consistently yield the same aberrant value for isotopic fractional abundance, it is likely they correspond to fragments that misidentified or subject to interference from overlapping signals).

In order to remove some of the outliers, fragments with consistently unreliable ^{13}C enrichment patterns were removed from the dataset. The data were separated into four different groups depending on whether it was derived from GFP or total protein samples and whether the seedlings had been grown without isotopic labelling or in 20% [$^{13}C_6$]-glucose. These datasets were then analysed separately for the four sample types: total protein from seedlings grown in normal glucose (Figure 24A), total protein from seedlings grown in 20% [$^{13}C_6$]-glucose (Figure 24 B), GFP purified from seedlings grown in normal glucose (Figure 25 A) and GFP purified from seedlings grown in 20% [$^{13}C_6$]-glucose (Figure 25 B). For each sample type, the fractional abundances of ^{13}C in approximately 500 amino acid fragments were compared with the total ion count.

All four datasets show a similar relationship between fractional abundance and total ion counts. At low ion counts, the data does not seem to faithfully report on the expected fractional abundance. This is not affected by the different protocols used to obtain the GFP measurements. Approximate 20% levels are reached at lower total ion counts than the expected levels are reached in corresponding samples with natural abundance of ^{13}C. This may be because of the difficulty of accurately

distinguishing between the low levels of ^{13}C enrichment and baseline noise in samples from material grown in unlabelled glucose in which ^{13}C is present at natural abundance.

Since the levels required for consistent detection of different amino acids and their fragments vary even when analysing unlabelled amino acid standards, it may not be practical to define a threshold value of the total ion counts required for accurate assessment of isotopic abundance. However, the aim of metabolic flux analysis experiments must be to reliably measure all types of labelling, including low levels of ^{13}C, and even when there is no prior information on the predicted levels of average isotopic abundance. The results from the comparison of isotopomer distributions and total ion count suggest that to achieve this, the total ion counts may have to be in the range of $1x10^5$ or greater.

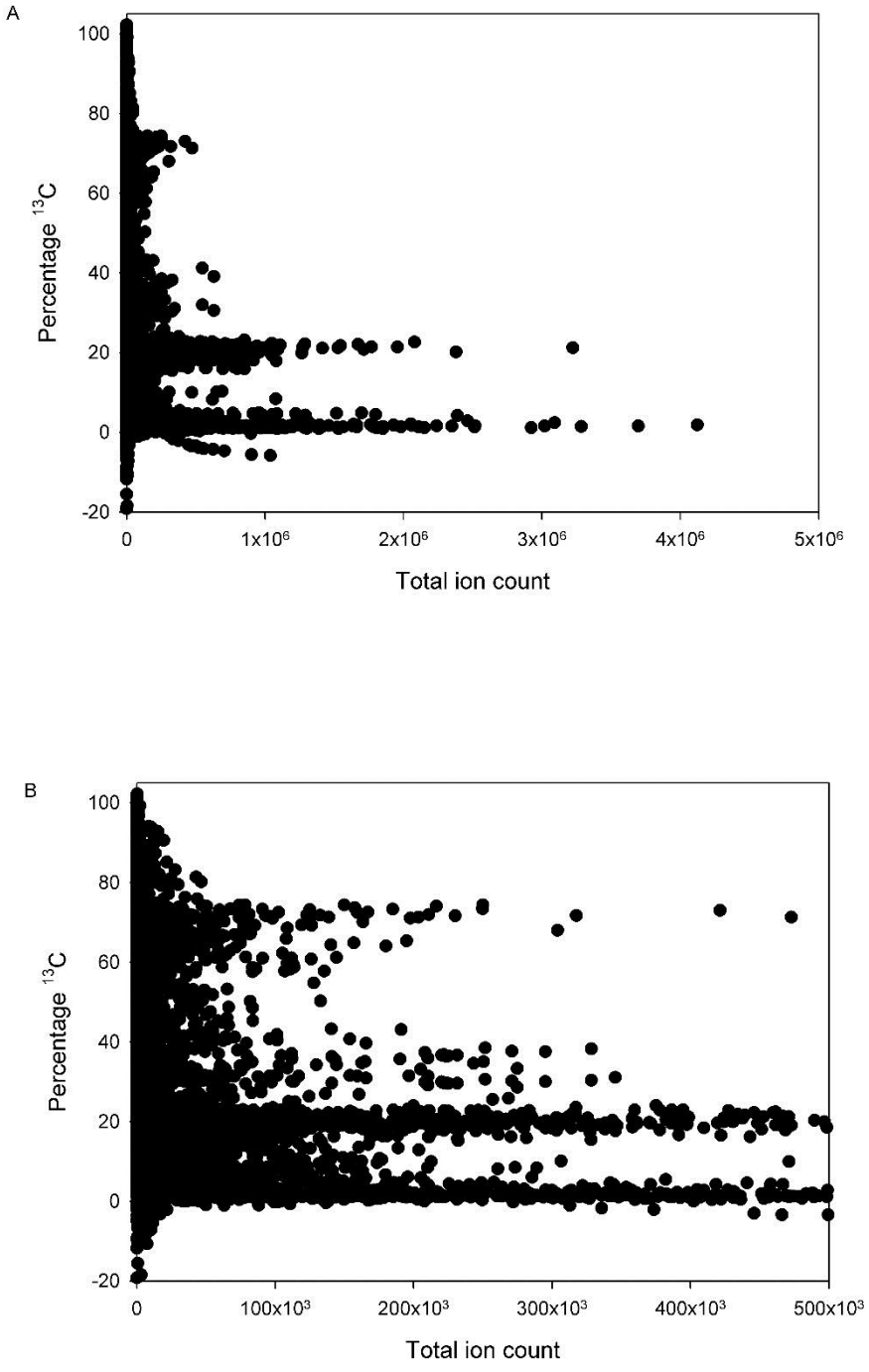


Figure 23: **Effect of abundance of an ion fragment on detection of the expected labelling pattern.** A dataset consisting of measurements from hydrolysed protein from seedlings grown with natural, 20%, or 80% [$^{13}C_6$]-glucose was used to plot the total ion counts and detected labelling. The dataset contained measurements from purified GFP and total protein fractions and were taken both from lines with GFP expression and the wildtype seedlings. (A) All fragments are shown. (B) Only fragments detected at total ion counts at or below 500,000 are shown.

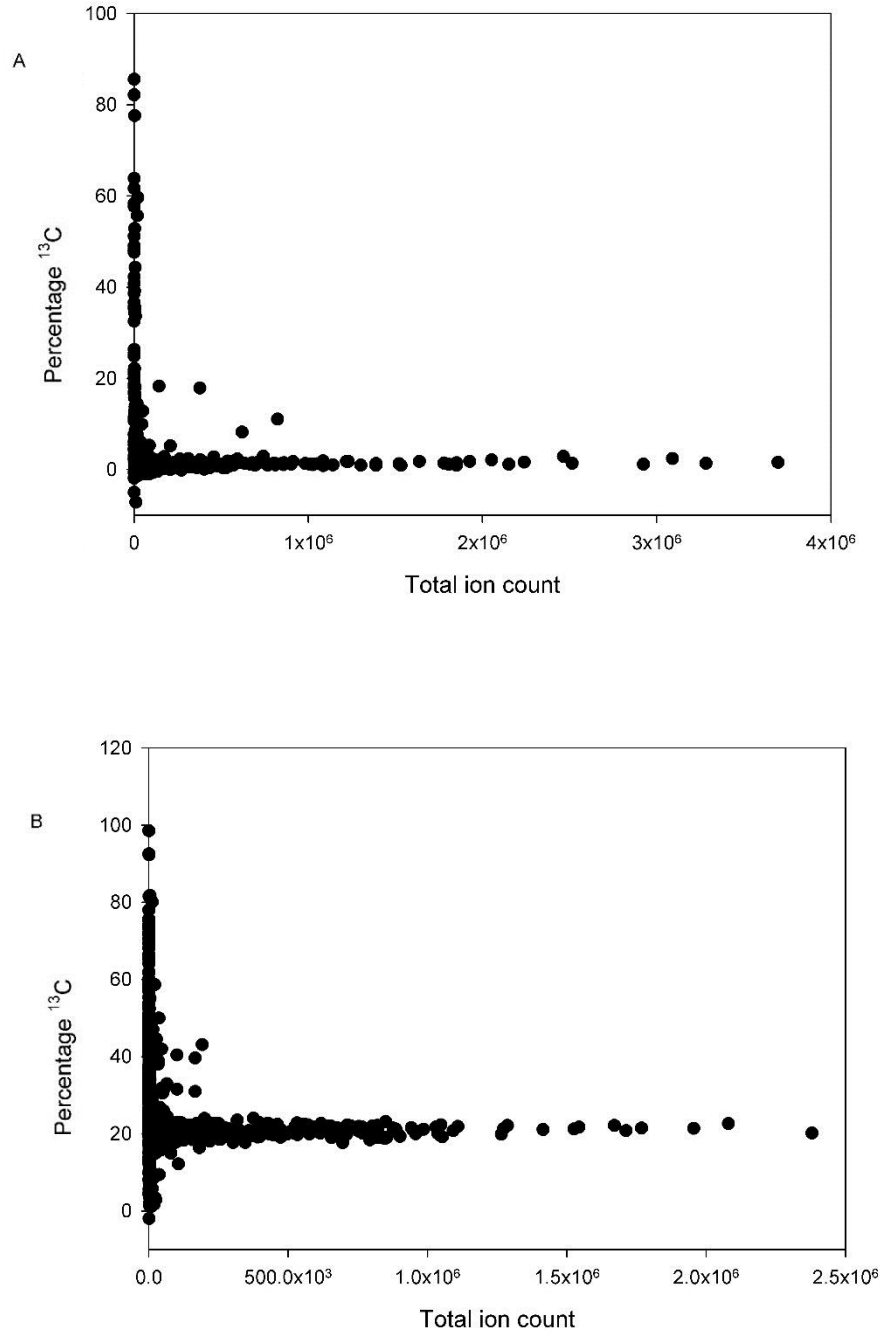


Figure 24: **Effect of abundance of an ion fragment on detection of the expected labelling pattern of measureable fragments derived from total protein.** A dataset consisting of measurements from hydrolysed protein from the total protein fractions of wildtype seedlings and seedlings with GFP expression was used to compare the total ion counts and average ^{13}C fractional abundance of fragments for which the labelling pattern could be reliably detected. (A) Protein extracted from seedlings grown with natural D-glucose. (B) Protein extracted from seedlings grown with 20% [$^{13}C_6$]-glucose.

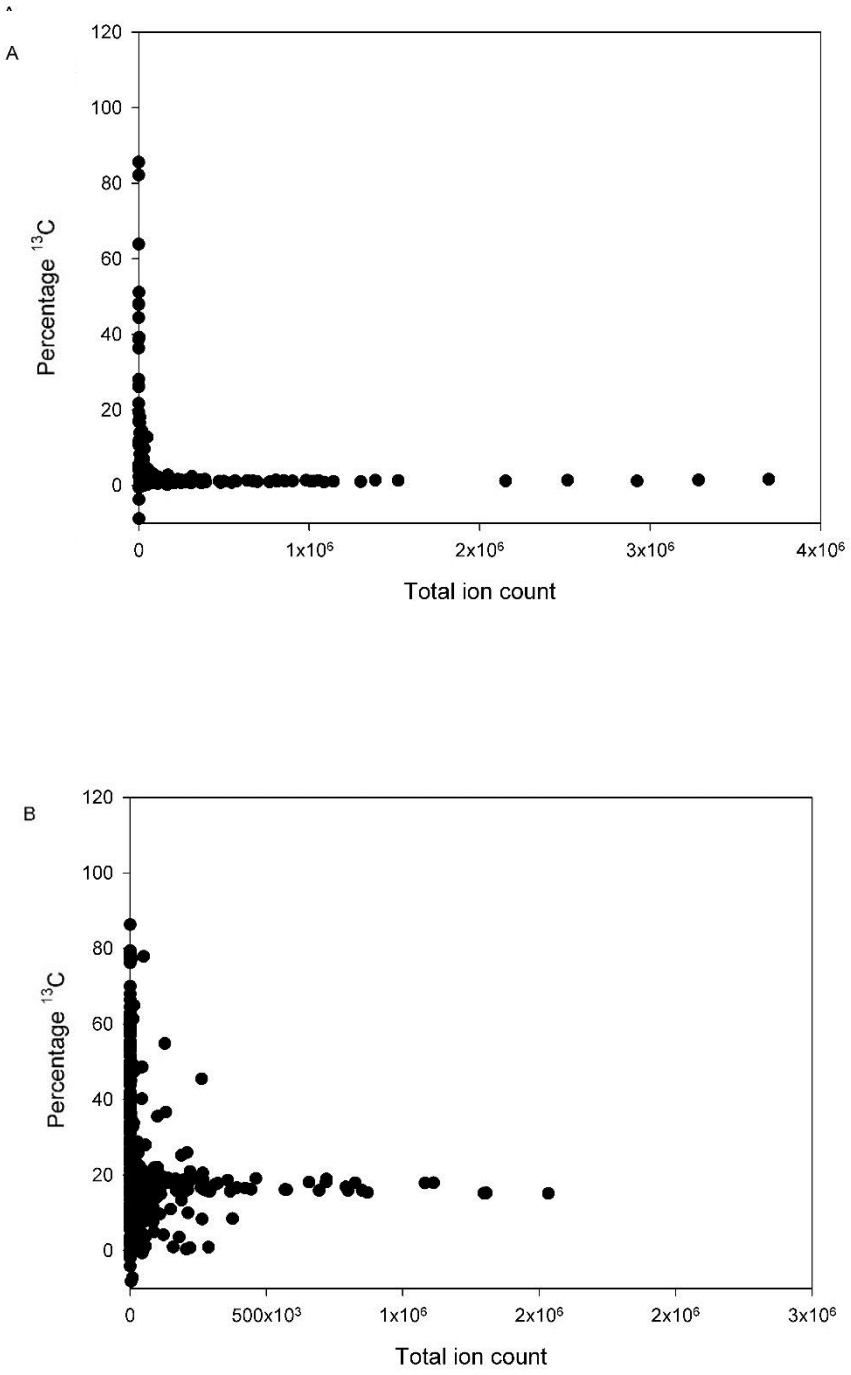


Figure 25: **Effect of abundance of an ion fragment on detection of the expected labelling pattern of measureable fragments derived from total protein.** A dataset consisting of measurements from hydrolysed immunopurified GFP from seedlings with constitutive GFP expression was used to compare the total ion counts and average ^{13}C fractional abundance of fragments for which the labelling pattern could be reliably detected. Protein extracted from seedlings grown with natural D-glucose. (B) Protein extracted from seedlings grown with 20% [$^{13}C_6$]-glucose.

4.2.3.2 Identification of amino acid fragments that provide reliable measurements of mass isotopomer distributions after hydrolysis

The relationship between total ion counts and fractional abundances was examined for individual amino acid fragments in the same way as described for the whole datasets in the previous section (Appendix 12). This enabled identification of a set of fragments that could be reliably measured from protein extracted from Arabidopsis seedlings. Some of the fragments derived from alanine (M-85 fragment) and glycine (M-85 and M-159 fragment) did not appear to produce reliable labelling data. In addition, some of the amino acids were not consistently detected at high enough levels to justify drawing conclusions from this type of analysis (histidine, methionine, lysine, arginine, proline, tyrosine and possibly leucine). For these amino acids the detection of the fragments was not consistent and therefore only smaller number of data points were detected at high levels from these data (Appendix 13). For all amino acid fragments except for glycine M-15, leucine M-15, M-85, M-159 the analysis agrees with published work on reliable amino acid fragments derived from total protein fractions (Zamboni *et al*. 2009). It was not possible to obtain consistent data on proline, methionine, histidine, tyrosine, and arginine. Table 6 summarises the amino acid fragments that were considered to be reliably detected from the datasets consisting of measurement data obtained both from GFP and total protein fractions.

For a small number of amino acids, some of the fragments that arise during the ionisation have overlapping masses. If two fragments that are made up of carbon atoms in different positions in the molecule have overlap in their mass peaks, it is impossible to distinguish between the ratio of the mass peak arising from each fragment in typical GC-MS data. Such fragments include alanine M-15 and f302, isoleucine and leucine M-57 and f302, and threonine M-159 and f302. These mass peaks should automatically be discarded. The selection of fragments for analysis has been described previously in published work (Antoniewicz et al., 2007; Dauner and Sauer, 2000; Zamboni et al., 2009).

Table 6: **Assessment of reliability of measurement of ^{13}C fractional abundance in protein-derived amino acids.** Data was collated based on comparison of total ion counts and fractional abundances for each individual amino acid fragment (Appendix 12). The fragments which could be reliably detected are indicated with X in the table. Where the fractional abundance at high levels differs from the predicted (based on the known isotopic labelling of the substrate), fragments are indicated as unreliable. Reasons for unreliable patterns are given where known.

Amino acid fragment	Reliable detection of fractional abundance	Reason for unreliable detection, if known	Sample type specific
Alanine M-15		Peak overlap	
Alanine M-57	X		
Alanine M-85		Unknown, fails quality tests in analysis (Zamboni *et al.* 2009)	Possibly – but similar results by Zamboni *et al*. 2009
Alanine M-159		Identified as instrument specific by Zamboni *et al*.	Possibly – but similar results by Zamboni *et al*. 2009
Alanine f302		Peak overlap	
Glycine M-15		Too low levels detected	Experiment specific
Glycine M-57	X		
Glycine M-85		Unknown, fails quality tests in analysis (Zamboni *et al.* 2009)	Possibly – but similar results by Zamboni *et al*. 2009
Glycine M-159		Identified as instrument specific by Zamboni *et al*.	Possibly – but similar results by Zamboni *et al*. 2009
Valine M-15	X		
Valine M-57	X		
Valine M-85	X		
Valine M-159	X		
Valine f302	X		
Leucine M-15		Low abundance	Experiment specific
Leucine M-57		Peak overlap	
Leucine M-85		Low abundance	Experiment specific
Leucine M-159		Low abundance	Experiment specific
Leucine f302		Peak overlap	

Amino acid fragment	Reliable detection of fractional abundance	Reason for unreliable detection, if known	Sample type specific
Isoleucine M-15	X		
Isoleucine M-57		Peak overlap	
Isoleucine M-85	X		
Isoleucine M-159	X		
Isoleucine f302		Peak overlap	
Serine M-15	X		
Serine M-57	X		
Serine M-85	X		
Serine M-159	X		
Serine f302	X		
Threonine M-15	X		
Threonine M-57	X		
Threonine M-85	X		
Threonine M-159		Peak overlap	
Threonine f302		Peak overlap	
Phenylalanine M-15	X		
Phenylalanine M-57	X		
Phenylalanine M-85	X		
Phenylalanine M-159	X		
Phenylalanine f302	X		

Amino acid fragment	Reliable detection of fractional abundance	Reason for unreliable detection, if known	Sample type specific
Aspartate M-15	X		
Aspartate M-57	X		
Aspartate M-85	X		
Aspartate M-159	X		
Aspartate f302	X		
Glutamate M-15	X		
Glutamate M-57	X		
Glutamate M-85	X		
Glutamate M-159	X		
Glutamate f302	X		
Proline		Low abundance	Experiment specific
Methionine		Low abundance	Experiment specific
Histidine		Low abundance	Experiment specific
Tyrosine		Low abundance	Experiment specific
Lysine		Low abundance	Experiment specific
Arginine		Low abundance	Experiment specific
Glutamine		Converted to glutamate during acid hydrolysis	
Asparagine		Converted to aspartate during acid hydrolysis	
Cysteine		Lost in hydrolysis	
Tryptophan		Lost in hydrolysis	

4.2.4 Modification of the analytical methodology can improve the detection of amino acids and reliability of the corresponding mass isotopomer distributions

The optimisation of standard protocols to allow detection of amino acids derived from constitutively expressed plant GFP purified from seedling cultures is a promising start for developing a protocol for analysis of specific cell types. Ultimately the aim of the work is to expand the analysis from constitutively expressed GFP to that expressed in a specific cell type in the Arabidopsis root. Quantification by fluorescence and the characteristic mass spectra obtained from GFP purified from the line with constitutive expression and the different cell types (Chapter 7) all indicate that the improvement achieved by the changes introduced into the protocol will not be sufficient to allow investigation of cell type specific metabolism in this system. Reliable detection from even lower amounts of amino acids might require the use of some other analytical technique.

In published work on reporter proteins the authors have suggested that large volume injections or more sensitive mass spectrometers could provide a solution for more consistent detection of isotope labelling (Allen et al., 2012; Rühl et al., 2011). To test the potential for improvement, two different mass spectrometry systems were compared to the standard Agilent 7890A with 5975c used for the work described previously in this chapter (and earlier in this thesis): a Trace GC Ultra from Thermo Fisher coupled to an ISQ mass spectrometer and Agilent 7890B with 7200 mass spectrometer. Both of these systems have the capability to accept large volume injections. For comparative purposes, the analytical protocol was kept as similar as possible for all three systems and, unless otherwise specified, samples were analysed in full scan mode.

4.2.4.1 Selective ion monitoring can improve detection of amino acids, but is not sufficient for detection of isotopomer distributions from the reporter protein

As described earlier in this chapter, SIM relies on detection of pre-defined masses or range of masses during specific times of the analysis and can allow significant increase in sensitivity of the analysis by increasing signal to noise ratio. This is achieved by using more of the scan time on detection of specific masses of interest. The signal to noise ratio or the ability to discriminate mass peaks from baseline noise is one of the key aspects of reliable detection of mass isotopomer distributions from small amounts of amino acids.

Before any modifications were introduced into the protocols, GFP samples were run using both SIM-mode and scan mode. The mass ranges and elution times chosen were those of the M-57 fragments corresponding to all the amino acids normally detected in GC-MS analysis. During the majority of runs, running in SIM mode did not result in any significant improvement in the sensitivity of detection of amino acid fragments (Figure 26). For some samples there was a significant increase in the total ion counts of amino acids obtained in SIM-mode compared to scan mode, but these remained at very low levels - in the range of tens of thousands. For the majority of samples, however, the total ion counts obtained using the two types of running modes were very similar. A similar lack of improvement when monitoring in SIM-mode was observed when higher levels of amino acids were detected due to changes in the protocol (Appendix 3 and Appendix 14). The increase in ion

counts observed as a result of SIM alone would not be sufficient to allow the analysis of mass isotopomer distributions from amino acids from the purified GFP. Applying SIM mode can therefore be a valuable addition to the toolbox for cell type specific MFA, but did not in this instance improve detection to the extent needed to have an appreciable impact on the reliability of the quantification of the isopotomeric profile of fragments that are present at low abundance.

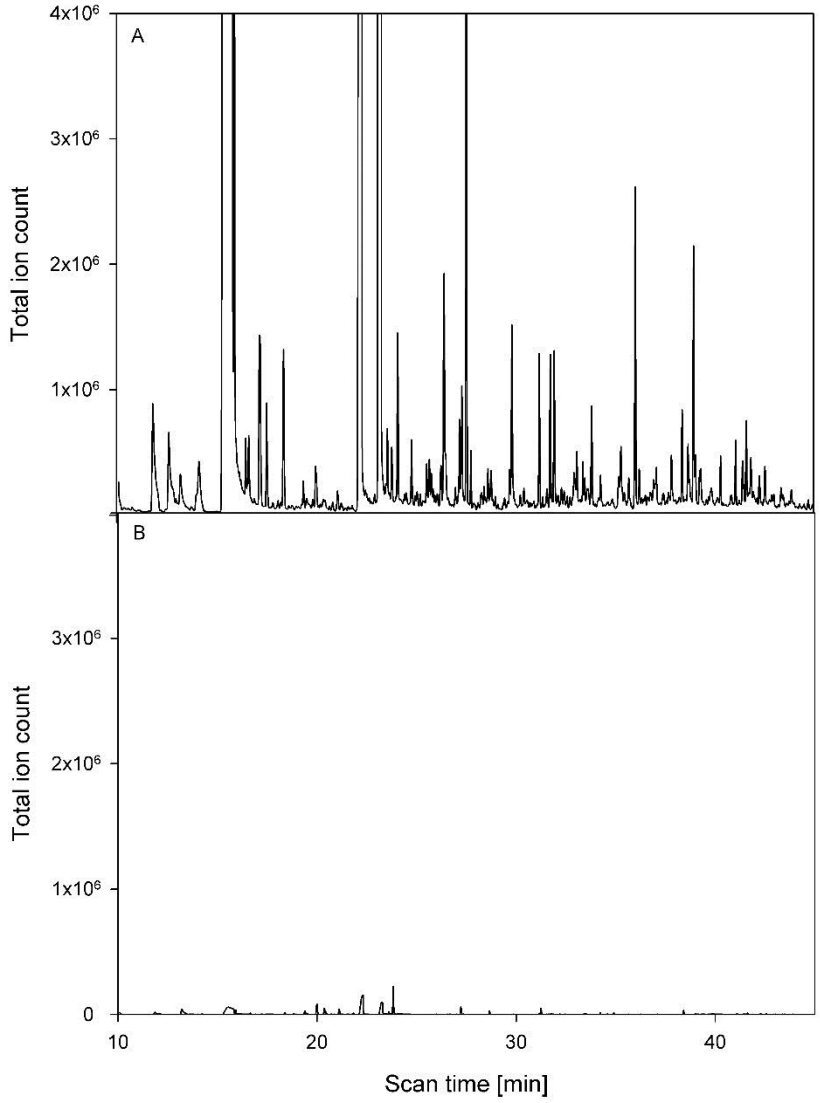


Figure 26: **Chromatographic analysis of amino acids derived from GFP using different detection modes.** GFP was immunopurified from 30-50 seedlings of Arabidopsis with constitutive GFP expression, hydrolysed and derivatised using conventional conditions. Aliquots of the same protein sample were analysed by GC-MS run in scan (A) and SIM (B) mode. SIM parameters were

set to measure the M-57 fragments of all the amino acids and elution times were identified from a GC-MS run in scan mode. The mass spectra for amino acid fragments detected in scan and SIM mode are presented in Appendix 3.

4.2.4.2 Large volume injections improve detectabilty by enabling analysis of greater quantities of amino acids

The injection volume for GC-MS experiments with standard inlets is mostly limited to approximately 2 µl due to the volume of solvent in the gas phase and the potential of liquid transfer back into the system. Therefore larger injection volumes normally require the use a specific multimode inlet. The advantage of using such systems is the ability to introduce more of the analyte into the system when the sample is dilute or when only a small amount of solution available and thus the use of split injections is less desirable. In order to explore the potential of large volume injections, a 5 µl aliquot of a hydrolysed GFP sample at a concentration 500 ng/µl was injected into a Thermo Fisher system (fitted with a large volume injector) linked to an ISQ mass spectrometer (Figure 27).

The amount of amino acids injected was equivalent to that of an entire batch of GFP purified from five cultures of 30-50 Arabidopsis seedlings with constitutive expression. The chromatographic analysis resulted in detection of most amino acids at total ion count levels in the range of approximately 10^7.This level of signal suggests the detection is not inherently more sensitive, but detection is improved because the system enables greater quantities of amino acids to be analysed (Figure 27). The experiment indicates of the potential levels of detectability that can be achieved using a relatively small number of GFP-expressing plant cultures.

When quantified with fluorescence, the levels of GFP in the lines with cell type specific expression were between 10 and 600 ng per culture. Even at the lowest levels of expression, injecting the whole sample purified from five cultures would result in total ion counts in the range of 10^5 (assuming a linear relationship between amount of protein and total ion count – see Appendix 6). Inlets capable of large volume injection can be installed on many standard GC-MS instruments and the main

limitation for how much they can improve the detection of low amounts of amino acids is the presence of other contaminants in the sample, as these may saturate the detector.

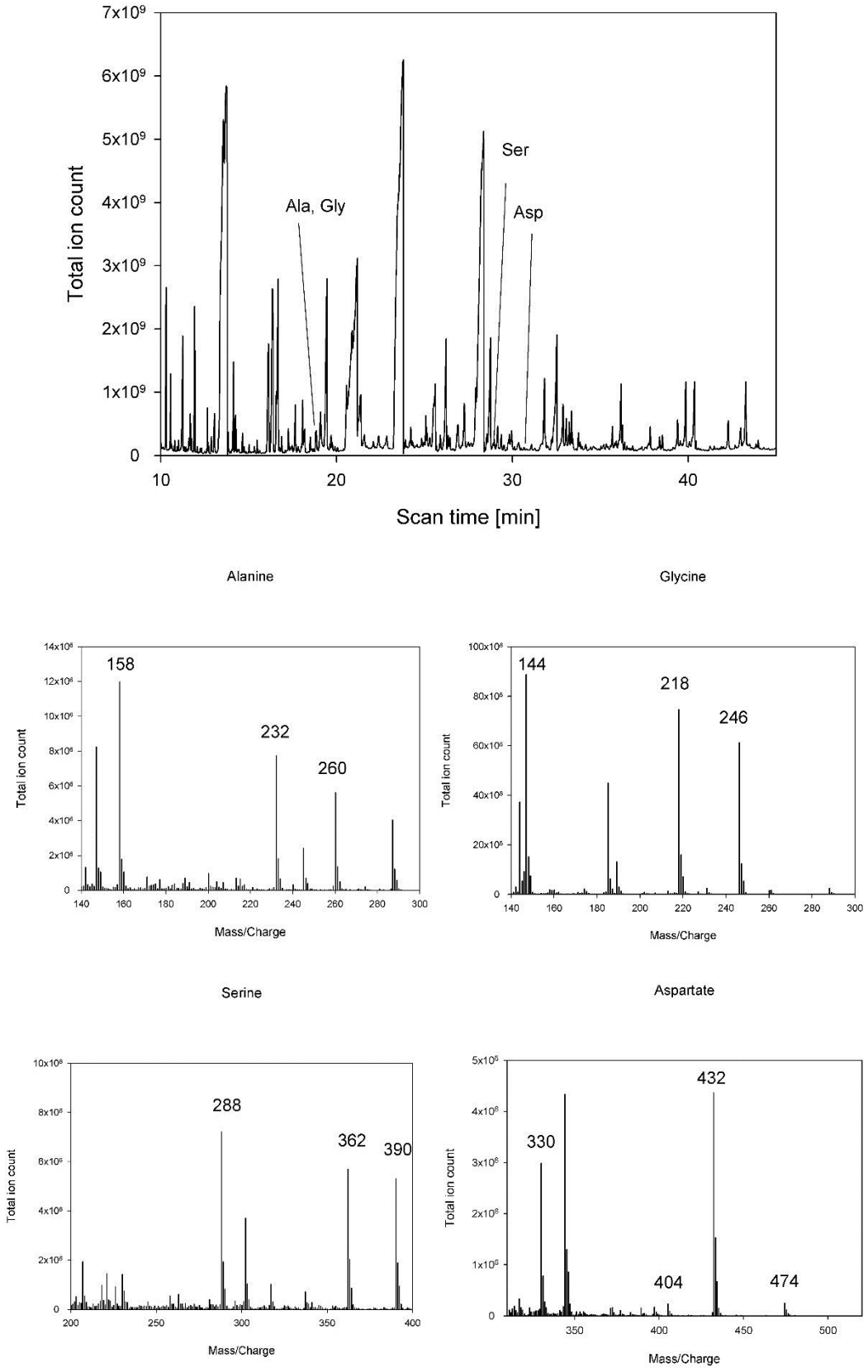


Figure 27: **Chromatographic analysis of a protein sample (500 ng/ µl) with Trace GC Ultra from Thermo Fisher with ISQ mass spectrometer using large volume injections (5 µl).**

4.2.4.3 Comparison between Agilent 5795C and Agilent 7200 instruments shows sensitivity of the instrument can affect the detection of amino acids

To test the suggestion that more sensitive mass spectrometers could help in detection of small amounts of reporter proteins, the Agilent 7200 mass spectrometer was used to analyse amino acids derived from a hydrolysed protein sample. Two samples of BSA of known concentrations (10 ng/µl and 20 ng/µl) were hydrolysed, and derivatised using conventional conditions. One aliquot of each sample was run using the Agilent 5975c system and another using the Agilent 7200 mass spectrometer. Injection volume, method and other conditions were kept the same for both systems and the samples were run simultaneously. The accurate mass capability of the Q-ToF was not explored in these experiments.

There was a significant difference between total ion counts detected by the 5975c and 7200 instruments (Figure 28). The total ion counts detected with the standard 5975c system (used for the analyses described earlier in this chapter) were consistent with the signal levels obtained from protein hydrolysates and the amino acid standards run earlier. The amino acids were detected in the order of 10^4 total ion counts from the 10 ng/µl standard and in the range of 10^5 from the 20 ng/µl standard. However, when an equivalent aliquot of the more dilute sample (10 ng/µl) was run with the 7200 instrument, the total ion counts were approximately 10^6. Across concentrations and amino acid fragments, the levels detected with the 7200 were significantly higher than those detected using the 5975c (Figure 28, Appendix 15). This difference in sensitivity was subsequently exploited when studying GFP from the cell type specific lines and is further demonstrated in Chapter 7. The improvement in detectability gained by use of a more sensitive instrument is specific to the instrument and methods used, but clearly offers a distinct advantage in the analysis.

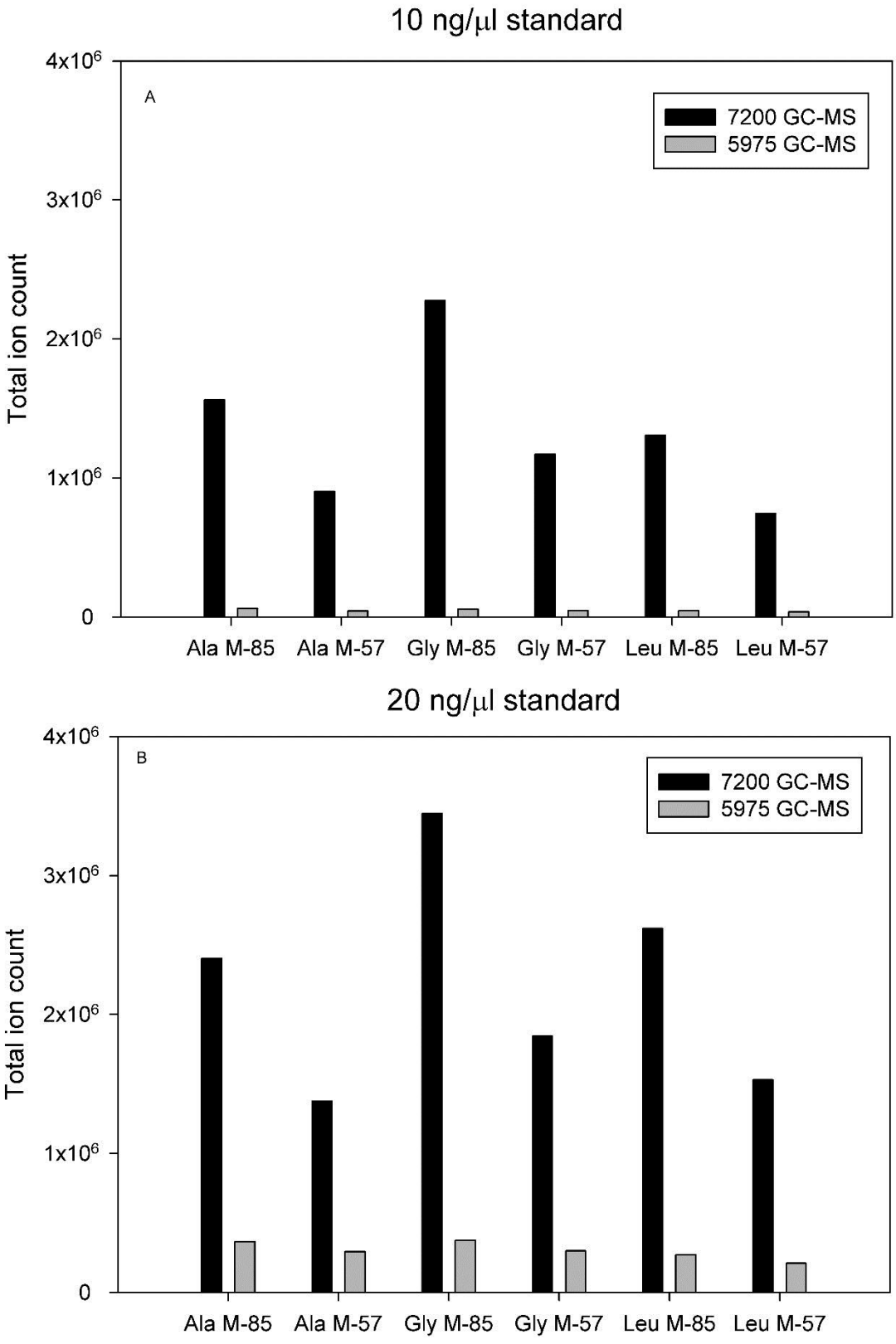


Figure 28: **Total ion counts for amino acid fragments detected using the Agilent 5975 and Agilent 7200.** Two samples of BSA standard solution with known concentrations were split into two aliquots, hydrolysed, dried, derivatised using standard conditions and run simultaneously with the

two Agilent instruments. (A) BSA standard with final concentration of 10 ng/µl analysed with Agilent 7200 (black) and Agilent 5975 (grey). (B) BSA standard with final concentration of 20 ng/µl analysed with Agilent 7200 (black) and Agilent 5975 (grey). Mass spectra of the individual amino acids are included in Appendix 15.

4.3 Discussion

The analytical requirements for a successful reporter protein strategy for metabolic flux analysis combine several factors: reducing contamination to a minimum, providing high levels of amino acids for analysis and using sufficiently sensitive detection techniques. This chapter has addressed each of these factors by introducing a highly specific purification of the reporter protein with the GFP nanotrap, reducing contamination by ultrafiltration and solid phase extraction, increasing the amounts of amino acids in the analysis, and finally by testing the effect of increased sensitivity on detection. Together these tests have made it possible to develop a protocol that can measure reliable mass isotopomer distributions from small amounts of constitutively expressed GFP from the Arabidopsis roots and provided the basis for smaller scale analysis that will be needed for GFP expressed in specific cell types.

By quantifying the different aspects of analysis, the feasibility of applying the reporter protein strategy in different systems can be assessed. The results suggest measurement of reliable isotopomer distributions requires signals far in excess of the minimum needed for identification and quantification of a metabolite and thus involves analysis of larger amounts of amino acids than would typically be required for conventional GC-MS analysis. In fact, detection of reliable mass isotopomer distributions may require up to 100 times higher total ion counts than were obtained for some amino acids derived from immunopurified GFP from the line with constitutive expression using conventional GC-MS protocols. Trying to achieve this by simply increasing the amount of biological material analysed seems highly impractical. In particular, given that even lower levels of GFP were purified from the lines with cell type specific expression, much more promising approach is to try to develop the standard protocols towards improving the sensitivity of detection of a greater proportion of the amino acids in a sample.

4.3.1 Amounts of amino acids sufficient for identification and quantitation by GC-MS may still be too low for reliable determination of mass isotopomer distributions

The results from the analysis of fractional abundances of ^{13}C and total ion counts show that in order to reliably measure all types of labelling, including natural abundance, the levels of amino acid fragments should ideally be at the levels of several hundreds of thousands. However, this does not mean that reliable data for some fragments cannot be measured at levels lower than this. In fact, in the comparison for individual amino acids, some labelling patterns seemed to be reliably measured at below 100,000 ion counts.

The difficulty with these measurements lies in defining when they are reliable, particularly if measurements of other fragments detected at similar signal intensities are showing significant variation. Moreover, in biological experiments using positionally labelled substrates (in which the average ^{13}C fractional abundance of metabolites cannot be predicted without a precise knowledge of the flux distribution), an alternative objective criterion must be found for whether the data accurately reports on the ^{13}C enrichment in the biological system and this is likely to be beyond the analytical instrument's capability of measuring it. In most MFA experiments, there is likely to be inherent (biological) variation in the redistribution of label between samples and so the fractional abundances themselves cannot be used to assess the analytical reliability of the measurements. More precise constraints can be set for data used in MFA experiments by using the error estimates provided by the mass correction tool MSCorr in mol% (Wahl et al., 2004). These estimates, however, are more useful for determining the reliability of a dataset of experimental data. They would not be able to report on what would be required to reduce the error.

However, the average fractional abundances from samples supplied with uniformly labelled substrate can be used to assess the reliability of isotopomer measurements of a particular fragment present at a particular abundance (total ion count). This allows us to be confident in exploiting measurements of the same fragment from a parallel sample supplied with a positionally labelled substrate, if it is detected at a similar level. The overall strategy would require the investigation to

always include a sample supplied with uniformly labelled substrate alongside replicate samples supplied with any positionally labelled substrate that was considered useful for flux determination in the specific network being investigated.

5 Validation of GFP as a reporter for ^{13}C labelling of total protein in Arabidopsis seedlings

Chapter 4 described the development of protocols for the reporter protein strategy and explored some of the analytical limits and how they can be overcome. The highly specific immunopurification of GFP from plant protein extracts using the GFP nanotrap was introduced to minimise contamination from biological and chemical sources and the factors affecting the detection of amino acids derived from small amounts of protein were considered and addressed in the analytical strategy. Chapter 5 focuses on the validation of the reporter protein strategy in *Arabidopsis thaliana* roots, which forms an integral part of the development of a method for investigating cell type specific metabolism in this system.

The detected mass isotopomer distributions are only useful for metabolic flux analysis, if they report accurately on the isotopomer distributions in the amino acid pools that are used for synthesis of total cellular protein. In order to confirm the GFP purified from the roots of *Arabidopsis thaliana* seedlings could be used as a reporter for cell type specific metabolism, the mass isotopomer distributions measured from amino acids derived from GFP were compared to those measured from the total protein fraction. Consistent with the data published on micro-organisms, Chapter 5 describes how both the average fractional abundance of ^{13}C calculated for the amino acid fragments in these fractions and the detected isotopomer distributions are indistinguishable between the reporter protein and the total protein fraction of *Arabidopsis thaliana* roots with constitutive GFP expression.

Another requirement for the use of GFP for the reporter protein strategy is that the expression of the foreign reporter protein does not perturb central carbon metabolism of the cells. To investigate any significant effects, the mass isotopomer distributions measured from amino acids derived from the total protein fraction from roots of Arabidopsis seedlings with constitutive GFP expression were compared to those in amino acids from wildtype ecotype Columbia seedlings grown under the same conditions. Separate analyses of isotopomer distributions and average fractional abundances of ^{13}C in amino acid fragments from the two total protein fractions show no significant variation caused by the GFP expression.

Finally, this chapter validates the use of hydroponically grown seedlings for metabolic flux analysis by exploring how the experimental system may be affecting the fractional abundances of ^{13}C that arise with 20% [$^{13}C_6$]-glucose as the substrate. The ^{13}C enrichment detected from protein in root and leaf fractions is approximately 20% in both cases, showing no significant differences between the data. As assimilation of atmospheric CO_2 would dilute the labelling detected from the amino acid fragments, there seems to be such a minor contribution from photosynthesis to the system that labelled glucose is being used as the main metabolic substrate under the experimental conditions used.

5.1 Introduction

The main requirement for the use of a reporter protein strategy for metabolic flux analysis is that it is able to provide data that are representative of the isotopomer distributions in amino acid fragments that would normally be derived from the total protein fraction. This can be expected to be the case if the component amino acids of both protein fractions arise from the same precursor pool of amino acids within the cell and analytical requirements for their consistent detection are met. The previous chapter focused onT showing how the use of a reporter protein strategy relies on the ability to detect the amino acid fragments in a reproducible manner from small amounts of purified protein. The focus of this chapter is on the reporter protein's capability to report on the intracellular fluxes by providing information on ^{13}C enrichment in amino acid fragments.

So far a reporter protein approach has been validated and used mainly in micro-organisms (Rühl et al., 2011; Shaikh et al., 2008; Shaikh et al., 2010), although there are two examples of work in plants where different highly abundant endogenous proteins were used. Allen and group used the small and large sub-units of RuBisCo purified from *Brassica napus* embryos (rapeseed) to report on subcellular differences in amino acid labelling (Allen et al., 2012). Since the large sub-unit is synthesised in the plastid and the small sub-unit in the cytosol, the amino acids derived from this protein could be used to show the two pools of amino acids are not isotopically equilibriated. In another example of the use of a reporter protein in plants, labelling in peptide fragments from storage proteins in soybean was tested as an alternative way to report on labelling in the total protein fraction (Mandy et al.,

2014). The strategy for detection of mass isotopomer distributions from peptides was suggested as a possible solution to problems with reliable GC-MS detection of a large set of amino acid fragments from small amounts of protein (Allen et al., 2014a; Mandy et al., 2014).

In both cases of work carried out in plant embryos, the authors reported potential limitations on application of their strategy. Allen and colleagues concluded that the first approach using GC-MS could only produce reliable data on a limited range of amino acid fragments (Allen et al., 2014e; Allen et al., 2012). On the other hand, the second technique using peptides might have limited use for metabolic flux analysis due to unreliable measurements, which the authors attributed to the analytical technique used (Allen et al., 2014a). The limited range of data arises from the limitations of GC-MS as an analytical technique. Hydrolysis conditions and detection limits affect the number of amino acid fragments that can be reliably detected from protein hydrolysates and the small amounts of protein used for approaches relying on reporter proteins challenge the detection limits of the techniques.

The protocols that are based on detection of isotopic labelling from peptide fragments may, on the other hand, suffer from limitations arising from the collisional fragmentation required for LC-MS/MS detection (Allen et al., 2014a). The same method has, however, been previously used in work with micro-organisms (Marco-Urrea et al., 2012) and similar aspects of the analytical technique have been tested for detection of other labelled metabolites (Rühl et al., 2012). These examples highlight the issues related to transferring methods that have been validated in other systems into work with plant tissues. In the case of a foreign reporter protein such as GFP, validating the strategy in the biological system for its ability to report on the isotopomer distributions in the amino acid pools of interest is therefore essential.

As previously discussed in Chapter 1, there are at least two important requirements that need to be met by a reporter protein strategy that aims to produce flux maps through steady state stable isotope MFA. First is the detection of a large enough set of measurements to determine the network of unique fluxes. Second is the requirement for isotopic and metabolic steady state and for describing all the

carbon transition reactions in the network accurately. The main difficulties could arise from assimilation of atmospheric carbon from CO_2 through photosynthesis. If such contributions are not included in the model, the label intake into the resulting network is poorly or inaccurately defined (Kruger et al., 2012; Roscher et al., 2000). In case of significant dilution of the isotopic labelling introduced in the substrate, it may be necessary to consider alternative labelling approaches for Arabidopsis seedlings (Heise et al., 2014). Both of these aspects can be investigated by a labelling approach using uniformly ^{13}C labelled substrate and measuring the fractional abundances of ^{13}C in fragment ions of amino acids. Assimilation of unlabelled carbon into the carbon skeletons used in central carbon metabolism would result in lower fractional abundances of ^{13}C, which could then be taken into account in models as a fluctuation in uptake rates (Kruger et al., 2012).

Surprisingly little information exists on the effects of GFP expression on the metabolic activity of cells. Expression of fluorescent proteins has been a widely used method to visualise location of proteins and other cell components in plants, and recently has become popular for experiments relying on cell sorting based on fluorescence to study gene expression and protein composition (for examples of such work in the Arabidopsis root see Birnbaum et al., 2003 and Lan et al,. 2013). While the popularity of the methods and the lack of reports of detrimental effects of GFP expression and suggest any such effects are unlikely to be drastic or physiologically important, they should not be simply ignored. Indeed, there may be significant effects for example in the studies on protein degradation (Yewdell et al., 2011) and over-expression of GFP has been reported to alter the physiology of micro-organisms (Allison and Sattenstall, 2007). Although it represents a different type of biological system, expression of recombinant proteins in yeast has been reported to cause changes in fluxes in central carbon metabolism (Jorda et al., 2012). As a foreign protein, GFP expression may burden the cells by affecting processes that are closely linked to the metabolic network under investigation, for example through protein synthesis or degradation, even if it has no clear function that would directly interfere with metabolic processes. The potential effect of GFP expression on central carbon metabolism should therefore be investigated as part of the validation of the strategy.

The validation of the reporter protein strategy for both feasibility and reproducibility are important to allow its use for future experiments that aim to study metabolism with a reporter protein and these aspects can differ in different plant systems and different growth conditions. The aim of this chapter is to validate the use of GFP as a reporter protein in the *Arabidopsis thaliana* roots using hydroponic cultures where [^{13}C]-glucose is provided as substrate. The specific objectives are 1) to examine the extent to which GFP reports on the labelling patterns in the pools of amino acids used for general protein synthesis, 2) to investigate the extent to which expression of recombinant GFP perturbs metabolism, and 3) the degree to which assimilation of atmospheric CO2 contributes to metabolism in the hydroponic cultures.

5.1.1 Analysis of mass isotopomer distributions obtained from stable isotope labelled biological tissue

When seedlings are grown in a mixture of [$^{13}C_6$]-glucose and D-glucose at natural abundance, the fractional abundance of ^{13}C in amino acid fragments derived from the total protein fraction will only differ from the fractional abundance of these substrates, if there is another unlabelled source of carbon (Masakapalli et al., 2013). Here the seedlings have been grown in the liquid medium containing the labelled glucose since germination and it is assumed that the contribution of pre-existing metabolite pools to biomass by the time tissue is harvested at 14 days is negligible. These assumptions underpin the criteria used across this chapter to assess fractional abundances of ^{13}C obtained from biological tissue: when 20% [$^{13}C_6$]-glucose is provided as a substrate, the expected ^{13}C enrichment of the biological tissue is equally approximately 20%. Deviation from this may arise as a result of the contribution from carboxylases that contribute to metabolic reactions in the cells. The use of atmospheric CO_2 by these enzymes results in incorporation of unlabelled carbon into intermediates of the TCA cycle, thus decreasing the fractional abundance of ^{13}C (Schwender and Ohlrogge, 2002).

The information imprinted on carbon skeletons of amino acids as they are processed through the biosynthetic network in a labelling experiment is measured in spectra as mass peaks that are processed into mass isotopomer distributions. Mass isotopomer distribution data contains the

information on the labelling pattern of the amino acids, but is difficult to compare without further processing. Normally metabolic flux analysis experiments rely on production of flux maps based on the metabolic network and the measured data. While they provide information on which data may not fit well into the model, these are complex representations of the metabolic phenotype and not ideal for assessing the quality of data or whether the isotope label is being incorporated from the substrate into metabolites of interest. This is especially the case, when the whole dataset needs to be validated.

One way to compare the data is to calculate the average ^{13}C enrichment in carbon atoms in the molecule, a method used for example to compare labelling in large and small subunits of RuBisCo by Allen and colleagues (Allen et al., 2012). This percentage represents the flow of labelled carbon into the target molecules and it is calculated directly from the mass isotopomer distributions. In the case of the reporter protein strategy it can be used to estimate how accurate and reproducible the detection of ^{13}C enrichment from a specific fragment is by comparing the values to those obtained from the total protein fraction. Detection of the ^{13}C enrichment from total protein extracts is highly reproducible as there are fewer limitations on the availability of protein, but nevertheless, this does not mean all amino acid fragments are reliably detected.

In Chapter 4 requirements were set for reliable detection of mass isotopomer distributions from small amounts of amino acids derived from hydrolysed protein, including detection at ion counts at or above 100,000. Although fractional abundances can be a good indication of the biological system reaching metabolic steady state, mass isotopomer distributions are the data ultimately used in the metabolic models. The focus of Chapter 5 is therefore on comparison of the mass isotopomer distributions obtained from the conventional measurements from amino acids derived from the total protein fraction of wildtype seedlings and those derived from the plants expressing GFP as well as the reporter protein itself. The mass isotopomer distributions for each molecule arise from using mass correction to adjust the raw data on total ion counts to compensate for the effect of the natural abundance of stable isotopes of other atoms in the native structure or in the chemical groups added during derivatisation (Wahl et al., 2004).

5.2 Results

5.2.1 Assessment of the reliability of detection of mass isotopomer distributions from amino acid fragments derived from GFP

In order to compare the stable isotope labelling introduced into amino acids in the total protein fraction and GFP in *Arabidopsis* plants with constitutive GFP expression, seedlings were grown for 14 days in liquid medium containing 20% [$^{13}C_6$]glucose and natural abundance D-glucose as substrate. At 14 days, roots and leaves of the seedlings were separated and protein was extracted from the roots. GFP was purified from three cultures of 30-50 seedlings using the immunopurification protocol and purification steps developed earlier in this study (Chapter 4). Both unfractionated protein and GFP samples were then hydrolysed into amino acids, derivatised with tBDMS and analysed with standard GC-MS instrument and method for detection of amino acid peaks. Lower volumes of derivatisation reagents were used for the GFP samples, as detailed in Chapter 4. The resulting mass-fragment data were processed to determine the isotopomer distributions and average ^{13}C fractional abundance in the fragments that arise through ionisation of the amino acid molecules.

Comparison of the average ^{13}C datasets for amino acid fragments derived from GFP and the total protein fraction revealed that the majority of amino acid fragments were reliably detected in samples from both sources. There are two main reasons for why all amino acids were not part of this group. Firstly, detection of some amino acids is more difficult in the reporter protein samples due to lower absolute levels of protein. Moreover, even if a particular amino acid is present at detectable levels, some fragments are generated less readily than others and may therefore not be measurable. Secondly, the procedures involved in purification of GFP result in the removal of contaminants that otherwise interfered with the detection of a particular fragment of a specific amino acid, such that fragments that were not consistently detected in total protein samples were detectable in GFP preparations. Additionally, some amino acids generated different, non-overlapping sets of fragments in sample from the two protein sources. Table 7 shows the amino acid fragments detected from only GFP and Table 8 shows the amino acid fragments detected only from the total protein fraction.

The labelling of most amino acids could be compared and a large dataset of data comparable to published work on reporter proteins was obtained. Exceptions included some important amino acids: histidine is the only amino acid derived from the ribose-5-phosphate precursor. It is often present at low levels in protein hydrolysates analysed by GC-MS. In these experiments, histidine fragments were detected reliably from the two types of samples, but different fragments were giving reliable measurements in the two sample types. Although in terms of metabolic flux analysis the fragments detected can provide information on the mass isotopomer distributions, they could not be directly compared as the fragments contain different carbon atoms of the molecule, as shown in Table 7 and Table 8.

Methionine and serine represent examples of amino acids that were reliably detected in only one sample type. Methionine is also one of the amino acids that are often present at lower levels and here it was only detected reliably from the purified GFP samples. Although low levels of methionine were present in some samples derived from the total protein fraction, the ion counts were too low to produce reliable estimates of the labelling patterns in a sufficient number of samples. Serine on the other hand was detected at high levels from both type of samples, but produced a reliable labelling pattern only in samples derived from the total protein fraction.

To systematically assess the suitability of specific amino acids for metabolic flux analysis the isotopomer distributions of each fragment were compared individually between the amino acids derived from purified GFP and those from the total protein fraction. The extent of equivalence of the labelling patterns of fragments from the two sources was assessed by linear regression analysis. For this analysis, glycine was omitted from the dataset due to the structure of the M-85 fragment, which consists of only one carbon that can be substituted with ^{13}C. Due to the small size of the carbon skeleton, the isotopomer distribution consists of two mass peaks (M+0 and M+1), which makes linear regression analysis for the specific fragment redundant.

For the rest of the amino acids, the isotopomer distribution for each amino acid fragment was compared by creating a matrix where the calculated average isotopomer distribution for each mass

peak of each fragment was paired with that from the other sample type. Figure 29 shows this comparison and the associated coefficient of determination.

Despite the good fit, the analysis reveals the isotopomer distribution of the arginine M-57 fragment as an outlier. Figure 29 B highlights the data points corresponding to the arginine isotopomer distribution and shows the graphical fit of isotopomer distributions between GFP and total protein fraction omitting the arginine data. Removing this data from the analysis increases the coefficient of determination to indicate a better fit of the trendline to data (0.995). The result is likely due to difficulties in reproducibly quantifying the isotopic labelling in arginine. In comparison of standard deviations of the isotopomer data, arginine values do show larger variation between all values, suggesting the detection of the peak is less reliable than for the other amino acids (Appendix 20).

To identify other fragments from GFP that may not report on the amino acid pool isotopomer distribution accurately, the isotopomer distributions for fragments derived from GFP were compared with those obtained from the total protein fraction individually for each of the amino acids (Figure 30). All comparisons resulted in coefficient of determination values of over 0.99 signifying a good correspondence, with the exception of arginine (already identified in the previous analysis) and the proline fragment M-85, which had a coefficient value of 0.984.

When the data provided for the M+0, M+1 and M+2 of the fragments obtained from GFP by Shaikh and colleagues was plotted against the total protein data, the coefficient of determination for the published dataset was over 0.99 showing a near perfect linear fit (Appendix 19). For amino acids the values of the coefficient of determination were similar to those obtained in this chapter, ranging from 0.995 for histidine to 1 for aspartate. This comparison suggests the amino acid isotopomer distribution data obtained from GFP purified from Arabidopsis seedlings is of similar quality to the data obtained and used in work with micro-organisms (Shaikh et al., 2008; Shaikh et al., 2010) and the additional demands on the analytical technique due to detection of a larger set of fragments and more mass peaks have not affected the data significantly. In addition, the quality of data is comparable to those obtained in previous flux analysis experiments using total protein data

(Masakapalli et al., 2013). Taking into account considerations detailed above regarding individual fragments where more variation may occur in the data, the isotopomer distributions obtained from purified GFP expressed in the Arabidopsis roots can be used to report on the isotopic labelling in the amino acid pool that is normally detected from the total protein fraction.

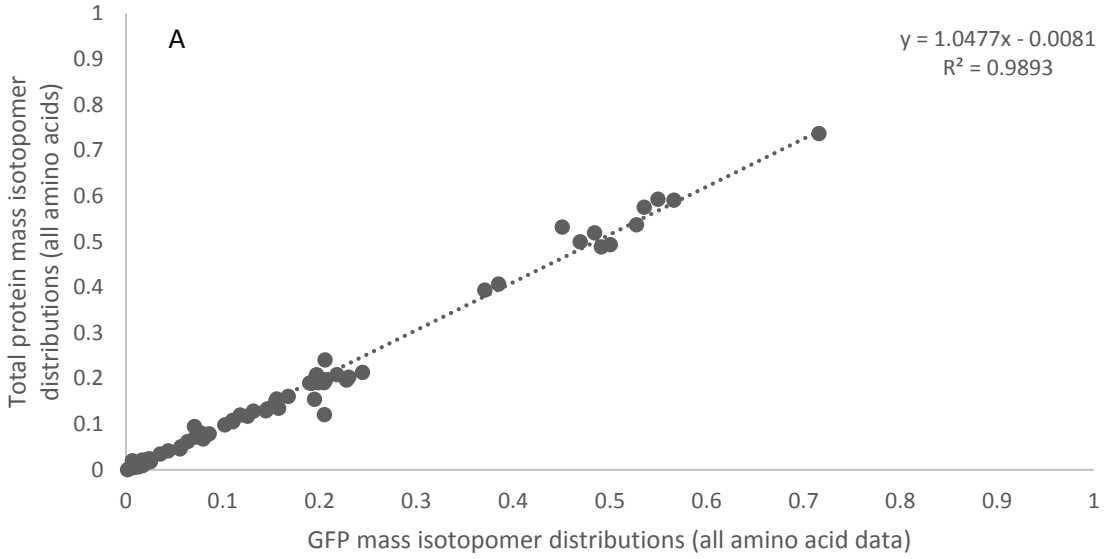
Table 7: **Average fractional abundances of ^{13}C (± standard deviations in percentage) for fragments detected reliably only from GFP samples with the corresponding structure of the fragments and precursor pathway shown.**

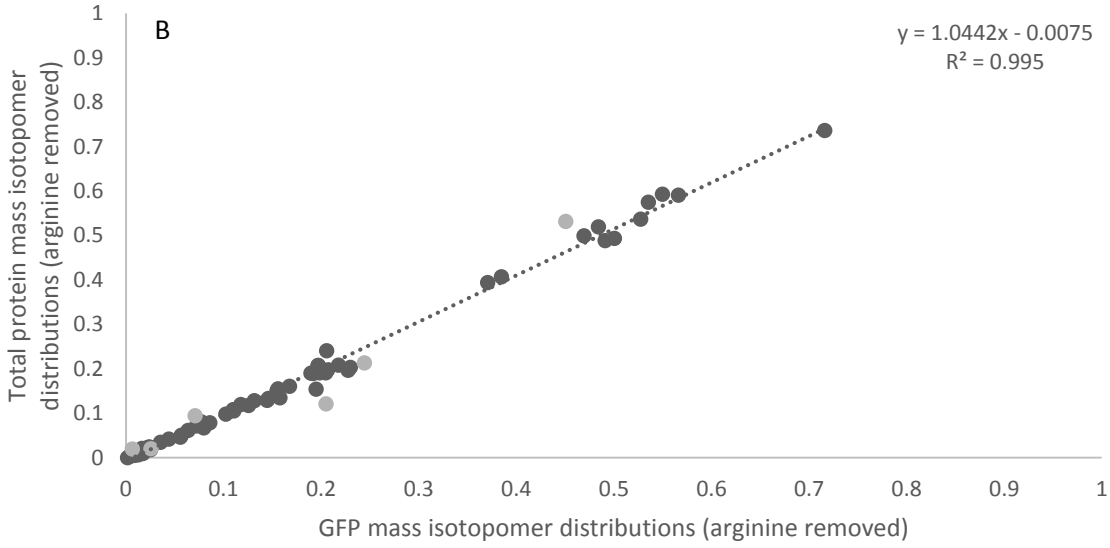
Derivatised molecule	Fragment ion	Carbon atoms	Precursor / pathway	Average ^{13}C fractional abundance (in %)
Alanine-2TBDMS	M-159	2,3	Pyruvate	19.47 ± 1.12
Glycine-2TBDMS	M-15	1,2	Serine	19.53 ± 1.24
Glycine-2TBDMS	M-159	2	Serine	20.45 ± 1.34
Valine-2TBDMS	f302	1,2	Pyruvate	18.66 ± 0.59
Leucine-2TBDMS	M-159	2,3,4,5,6	Pyruvate and Acetyl-CoA	20.04 ± 1.87
Methionine-2TBDMS	M-57	1,2,3,4,5	Aspartate	20.06 ±0 .55
Methionine-2TBDMS	M-85	2,3,4,5	Aspartate	19.56 ± 0.53
Methionine-2TBDMS	M-159	2,3,4,5	Aspartate	19.51 ± 1.13
Methionine-2TBDMS	f302	1,2	Aspartate	20.72 ± 2.91
Threonine-3TBDMS	M-159	2,3,4	Aspartate	20.17 ± 1.01
Glutamate-3TBDMS	f302	1,2	Alpha-ketoglutarate	19.64 ± 0.47
Histidine-3TBDMS	M-159	2,3,4,5,6	Ribose-5-P	18.77 ± 2.24
Histidine-3TBDMS	f302	1,2	Ribose-5-P	19.94 ± 2.04

Table 8: **Average fractional abundances of ^{13}C (± standard deviations in percentage) for fragments detected reliably only from samples derived from the total protein fraction with the corresponding structure of the fragments and precursor pathway shown.**

Derivatised molecule	Fragment ion	Carbon atoms	Precursor / pathway	Average fractional abundance of ^{13}C (in %)
Alanine-2TBDMS	f302	1,2	Pyruvate	18.99 ± 0.67
Glycine-2TBDMS	M-57	1,2	Serine	19.76 ± 0.72
Proline-2TBDMS	M-15	1,2,3,4,5	Glutamate	19.45 ± 0.15
Proline-2TBDMS	M-57	1,2,3,4,5	Glutamate	20.22 ± 0.93
Serine-3TBDMS	M-15	1,2,3	3-Phosphoglycerate	19.71 ± 1.04
Serine-3TBDMS	M-57	1,2,3	3-Phosphoglycerate	19.66 ± 0.90
Serine-3TBDMS	M-85	2,3	3-Phosphoglycerate	19.71 ± 0.91
Serine-3TBDMS	M-159	2,3	3-Phosphoglycerate	19.54 ± 0.98
Serine-3TBDMS	f302	1,2	3-Phosphoglycerate	19.99 ± 1.04
Aspartate-3TBDMS	M-15	1,2,3,4	Oxaloacetate	20.22 ± 0.96
Aspartate-3TBDMS	M-159	2,3,4	Oxaloacetate	20.42 ± 1.42
Glutamate-3TBDMS	M-15	1,2,3,4,5	Alpha-ketoglutarate	20.19 ± 1.14
Lysine-3TBDMS	M-85	2,3,4,5,6	Pyruvate	19.90 ± 0.25
Histidine-3TBDMS	M-15	1,2,3,4,5,6	Ribose-5-P	21.13 ± 0.40
Histidine-3TBDMS	M-57	1,2,3,4,5,6	Ribose-5-P	20.95 ± 0.63
Tyrosine-3TBDMS	M-15	1,2,3,4,5,6,7,8,9	P-enolpyruvate and erythrose-4-P	19.86 ± 1.18
Tyrosine-3TBDMS	M-57	1,2,3,4,5,6,7,8,9	P-enolpyruvate and erythrose-4-P	18.05 ± 1.69
Tyrosine-3TBDMS	M-85	2,3,4,5,6,7,8,9	P-enolpyruvate and erythrose-4-P	19.86 ± 0.33
Tyrosine-3TBDMS	f302	1,2	P-enolpyruvate and erythrose-4-P	19.40 ± 0.27

Figure 29: **Comparison of mass isotopomer distributions of amino acids derived from GFP and total protein.** Total ion counts measured from amino acid mass peaks derived from the total protein fraction and immunopurified GFP from seedlings grown in 20% [$^{13}C_6$]-glucose were processed into mass isotopomer distributions to obtain average data on each mass peak corresponding to each fragment. These data were compared between the total protein fraction and GFP by fitting a line through the data and calculating a coefficient of determination. A) Matrix containing mass isotopomer data for at least one fragment for all the amino acids that were detected, B) Arginine fragment M-57 mass isotopomer data is highlighted as an outlier and coefficient of determination calculated for the rest of the data. The value increases as arginine data is removed from the main dataset.





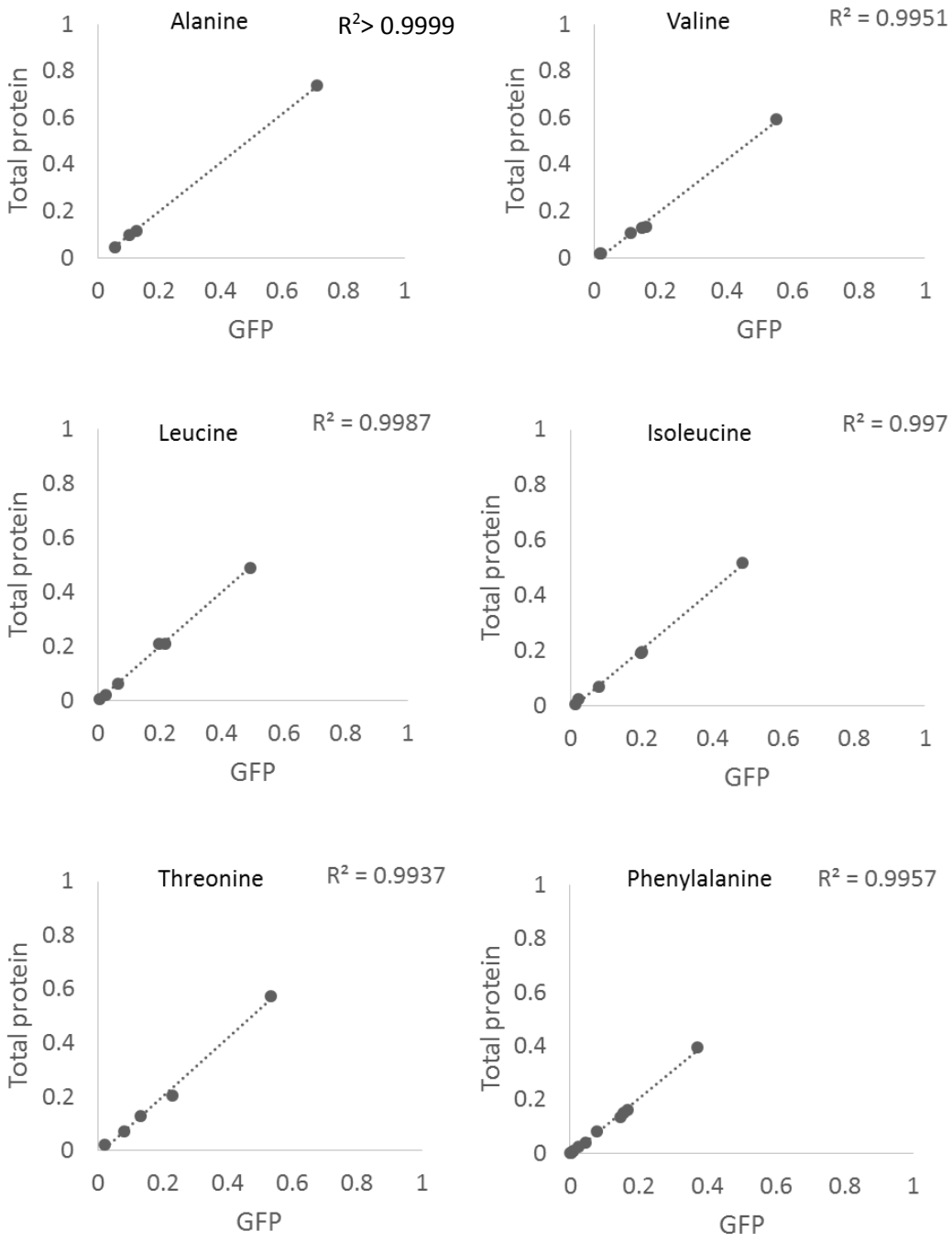
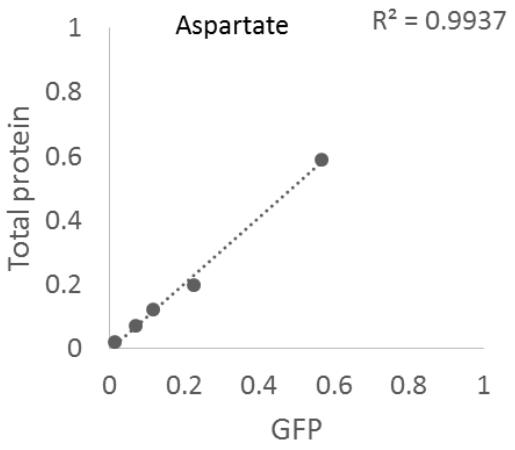
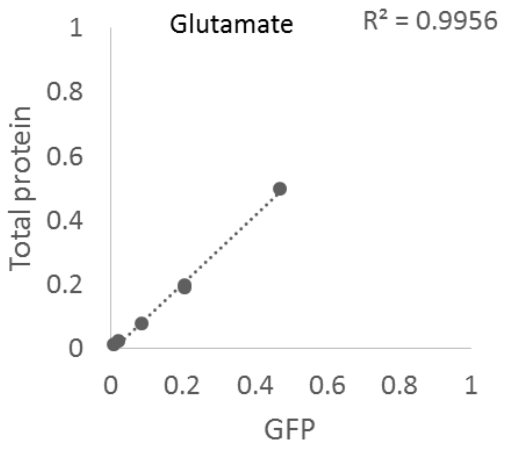
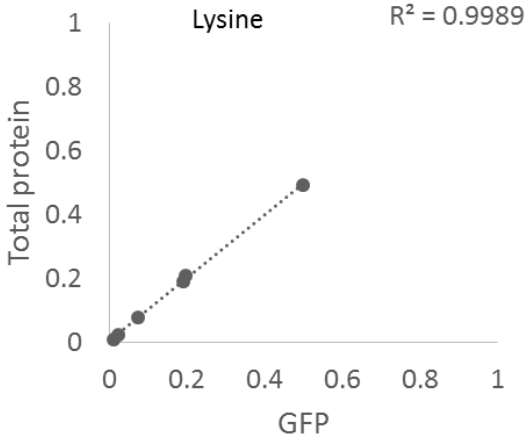
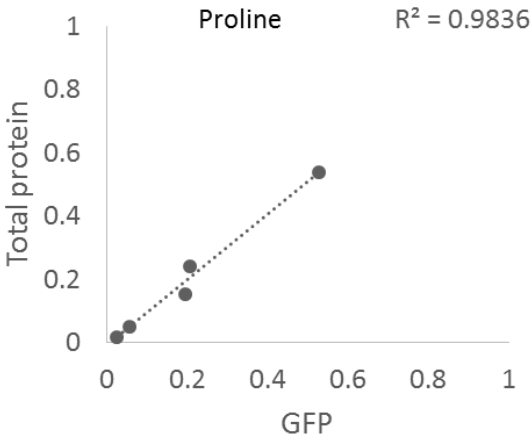


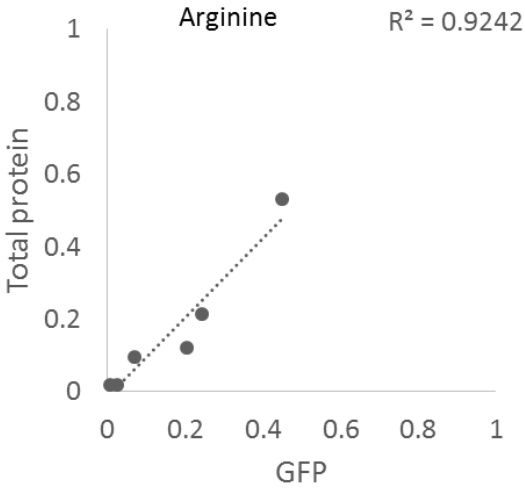
Figure 30: **Comparison of mass isotopomer distributions of individual amino acid fragments derived from GFP and total protein.** Total ion counts measured from amino acid mass peaks derived from the total protein fraction and immunopurified GFP from seedlings grown in 20% [$^{13}C_6$]-glucose were processed into mass isotopomer distributions to obtain average data on each mass peak corresponding to each fragment. These data were compared between the total protein fraction and GFP by fitting a line through the data and calculating a coefficient of determination for the amino acid fragments that were reliably detected across datasets.

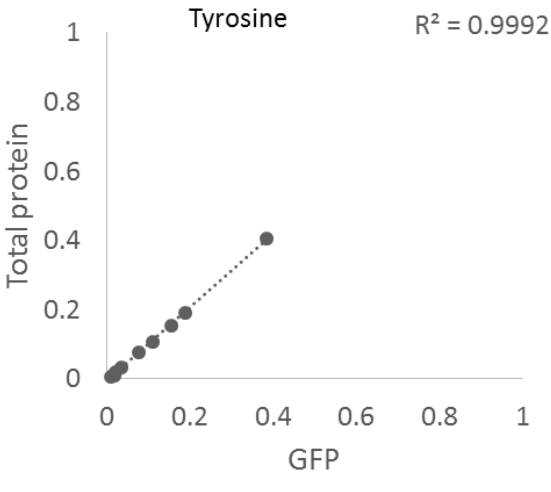

Aspartate
R² = 0.9937
Total protein
GFP
0
0.2
0.4
0.6
0.8
1


Glutamate
R² = 0.9956
Total protein
GFP
0
0.2
0.4
0.6
0.8
1


Lysine
R² = 0.9989
Total protein
GFP
0
0.2
0.4
0.6
0.8
1


Proline
R² = 0.9836
Total protein
GFP
0
0.2
0.4
0.6
0.8
1


Arginine
R² = 0.9242
Total protein
GFP
0
0.2
0.4
0.6
0.8
1


Tyrosine
R² = 0.9992
Total protein
GFP
0
0.2
0.4
0.6
0.8
1

5.2.2 GFP reports on the average fractional abundances of ^{13}C in amino acid fragments derived from total protein in Arabidopsis root

The mass isotopomer distributions for amino acid fragments derived from GFP and the whole root protein fraction were used to calculate the average fractional abundance of ^{13}C by taking into account the number of ^{13}C substitutions in each of the mass peaks and taking an average of the sum of these multiplied values, as described in Chapter 2. For GFP, 15 biological replicates were compared, although not all amino acid fragments were detected in all of the samples. For the total protein fraction 19 biological replicates were analysed. Only fragments detected in at least three biological replicates were included in the subsequent analysis.

The comparison between average ^{13}C percentage derived from GFP and from total protein is shown in Appendix 16. All of these amino acid fragments were labelled to approximately 20%, as expected, and there was no significant difference detected between the labelling detected in amino acid fragments derived from GFP and those derived from the total protein fraction. This result was further confirmed by statistical analysis of the two datasets with the Student's t-test, in which all comparisons yielded p-values above 0.08. The aspartate fragment showed lower average ^{13}C percentage in GFP samples, which may be an effect from poor detection. The ^{13}C enrichment therefore supports the conclusions drawn from the mass isotopomer data: the information derived from GFP immunopurified from Arabidopsis roots can be used to report on the isotopic labelling in the total protein fraction. In addition, the data suggests the seedlings had reached isotopic steady state in the hydroponic cultures.

A few amino acids formed an exception or could not be directly compared in the analysis. For histidine, the detected fragments provide data partly on different carbons, but both datasets suggest the fragment contains approximately 20% ^{13}C. For serine, the average ^{13}C fractional abundances in amino acids derived from GFP were lower than expected. The unreliable labelling was confirmed by the variation not only in the samples labelled with 20% [$^{13}C_6$]-glucose, but also those that were not labelled. The effect could cause the ^{13}C enrichment to consistently drop to approximately 15%

in labelled samples and below 0.5% in samples with no additional labelling with the amino acid fragments being detected at high abundances and was consistent across fragments. Based on the change in samples from cultures with no added labelled substrate, the effect was ascribed to peak overlap that affected the analytical detection of serine from GFP.

5.2.3 There are no significant differences in mass isotopomer distributions or fractional abundances of ^{13}C detected from leaf and root tissue of Arabidopsis seedlings grown in 20% [$^{13}C_6$]-glucose

The possible contribution from CO_2 uptake in a stable isotope labelling experiment in Arabidopsis seedlings grown in hydroponic cultures was investigated by extracting shoot and root protein fraction from seedlings grown in 20% [$^{13}C_6$]-glucose. The samples were analysed as described before for total protein samples by hydrolysing the protein into amino acids, which were derivatised with tBDMS and detected using GC-MS. The data was processed to derive mass isotopomer distributions for each amino acid fragment and the calculated average mass isotopomer distributions for each mass peak of each fragment was compared between the two sources (Figure 31).

In this analysis, the coefficients of determination for all amino acid fragments included in the analysis except for phenylalanine were over 0.99, indicating a good fit. For phenylalanine the coefficient was just over 0.94. The decrease in the coefficient of determination is likely to arise from unreliable detection of higher mass isotopomers of a particular fragment which are present at very low levels. Amino acids not detected consistently in both datasets were not included in the analysis.

The fractional abundances of ^{13}C were calculated from the mass isotopomer data and fragments with reliable isotopic labelling patterns were compared between different amino acid fragments (Appendix 17). The average fractional abundances of ^{13}C in the roots and leaves were approximately 20% for all the amino acid fragments. The Student's t-test concluded that there was no significant difference between labelling in the roots and leaves, with p-values above 0.13.

Based on these comparisons, the contribution of unlabelled carbon through CO_2 from photosynthesis was so minimal, that no significant difference to the mass isotopomer distributions or fractional abundance of ^{13}C in the original substrate was observed in roots or leaves. The earlier measurements from the root total protein fraction and GFP had indicated these levels of ^{13}C enrichment in the roots, but the experiment on leaves suggests CO_2 is making an equally negligible contribution to the carbon metabolism of leaves in this system. Green leaves were present on all seedlings and on visual inspection they had no obvious phenotypic defect that could explain the apparent lack of photosynthesis.

5.2.4 Over expression of GFP does not influence the ^{13}C labelling of cellular proteins in transgenic Arabidopsis seedlings

Total protein fractions were extracted from the wildtype Columbia seedlings and the GFP over expression line seedlings grown in 20% [$^{13}C_6$]-glucose. Protein was extracted from seedlings of both lines, analysed by GC-MS and the resulting data was processed. Labelling measured from the two sample types was compared in order to identify any effect on the pattern of redistribution of label from the supplied ^{13}C-substrate in response to the GFP expression. The data was processed to derive mass isotopomer distributions for each amino acid fragment and the calculated average mass isotopomer distributions for each mass peak of each fragment in the total protein fractions from wildtype seedlings and was compared with that from seedlings with GFP expression (Figure 32). In this comparison all of the coefficients of determination were over 0.99, corresponding to a good fit of the data.

The mass isotopomer distributions were used to calculate average fractional abundances of ^{13}C and these were compared between different amino acid fragments (Appendix 18). The ^{13}C enrichment in amino acid fragments was found to be at similar in protein from both plant lines (approximately 20%). The close similarity (equivalence) of the two sets data indicate that GFP expression has no significant effect on the redistribution of label within the system, and provides compelling evidence

that flux through the network of central carbon metabolism in roots is not appreciably affected by expression of GFP at the levels achieved by the plant lines used in this study.

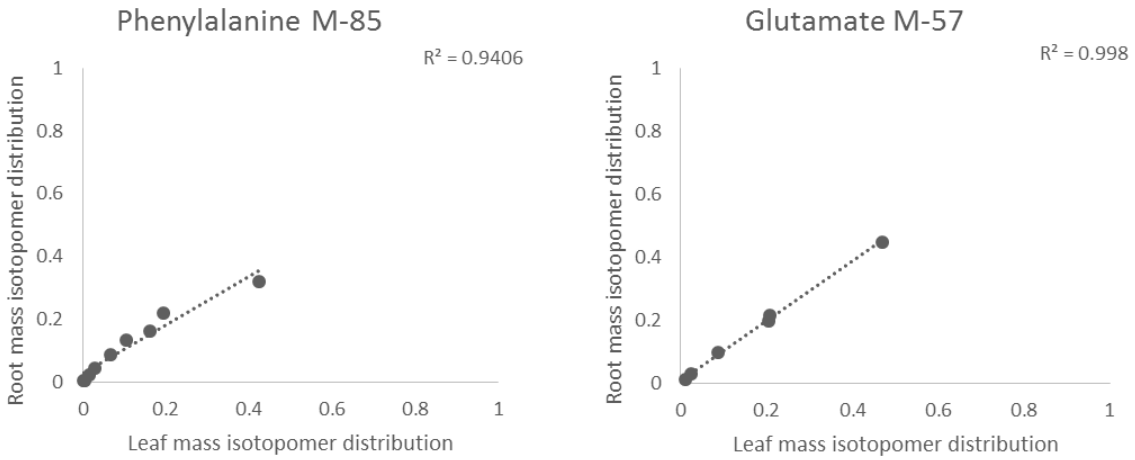
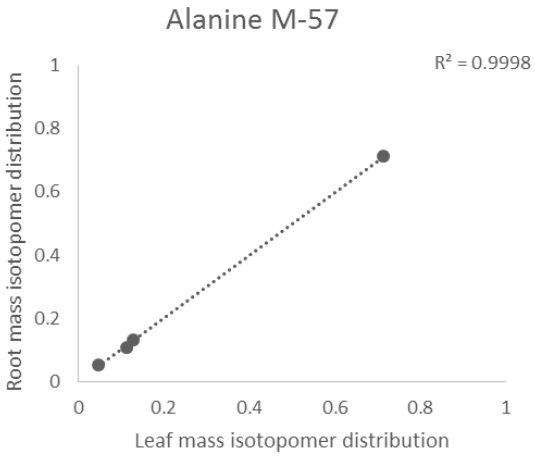
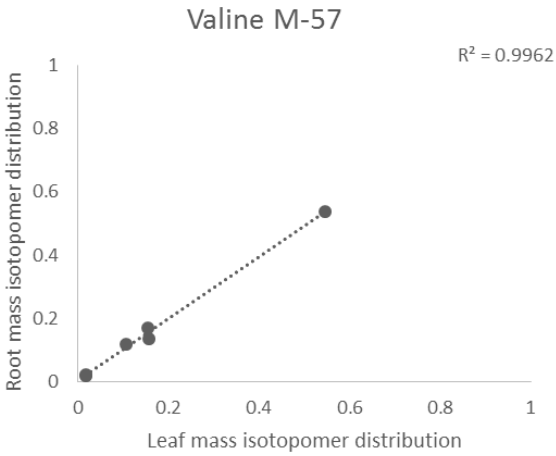
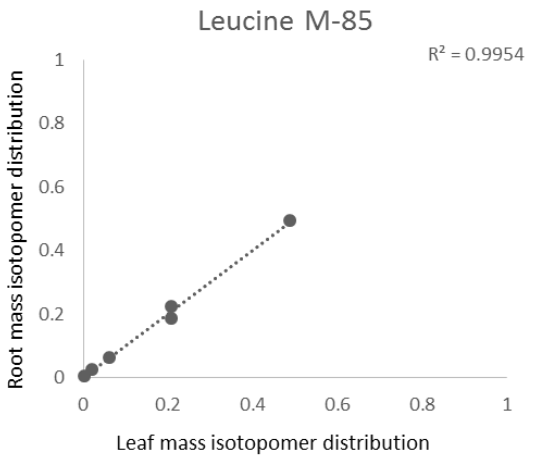
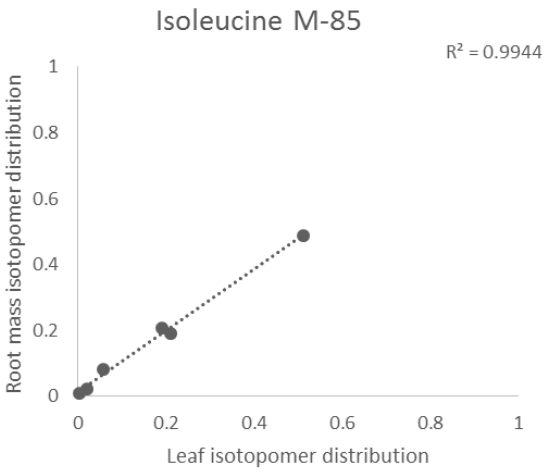


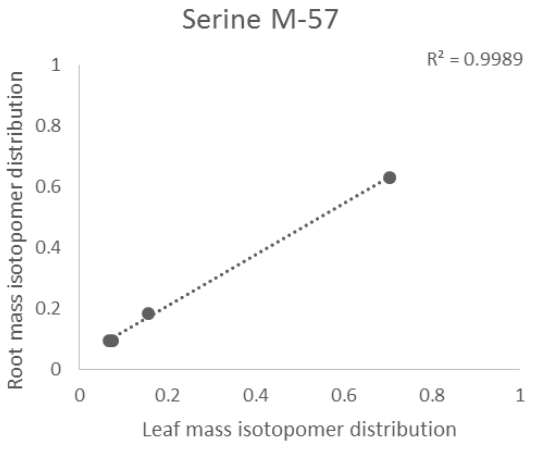
Figure 31: **Comparison of mass isotopomer distributions in amino acid fragments from the total protein fraction in leaves and roots of *Arabidopsis thaliana* seedlings.** Total ion counts measured from amino acid mass peaks derived from the total protein fraction from leaf and root tissue from seedlings grown in 20% [$^{13}C_6$]-glucose in hydroponic cultures were processed into mass isotopomer distributions to obtain average data on each mass peak corresponding to each fragment. These data were compared between the leaf and root tissue by fitting a line through the data and calculating a coefficient of determination.

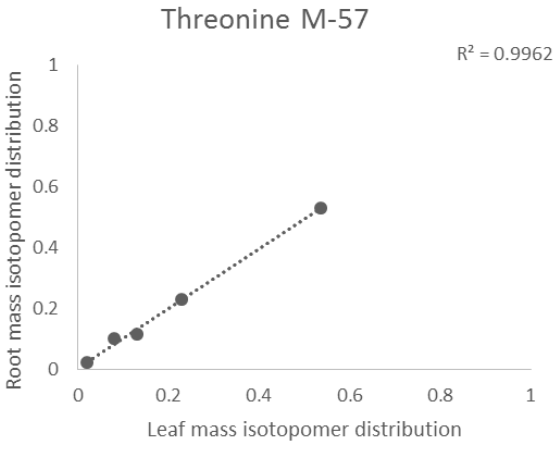

Alanine M-57
R² = 0.9998
Root mass isotopomer distribution
Leaf mass isotopomer distribution
0
0.2
0.4
0.6
0.8
1

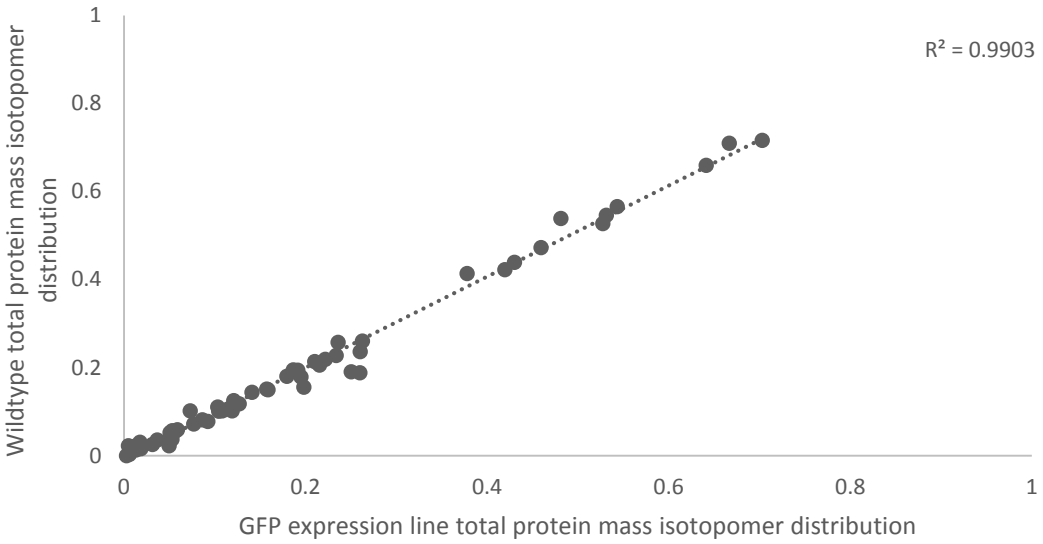

Valine M-57
R² = 0.9962
Root mass isotopomer distribution
Leaf mass isotopomer distribution
0
0.2
0.4
0.6
0.8
1


Leucine M-85
R² = 0.9954
Root mass isotopomer distribution
Leaf mass isotopomer distribution
0
0.2
0.4
0.6
0.8
1


Isoleucine M-85
R² = 0.9944
Root mass isotopomer distribution
Leaf isotopomer distribution
0
0.2
0.4
0.6
0.8
1


Serine M-57
R² = 0.9989
Root mass isotopomer distribution
Leaf mass isotopomer distribution
0
0.2
0.4
0.6
0.8
1


Threonine M-57
R² = 0.9962
Root mass isotopomer distribution
Leaf mass isotopomer distribution
0
0.2
0.4
0.6
0.8
1



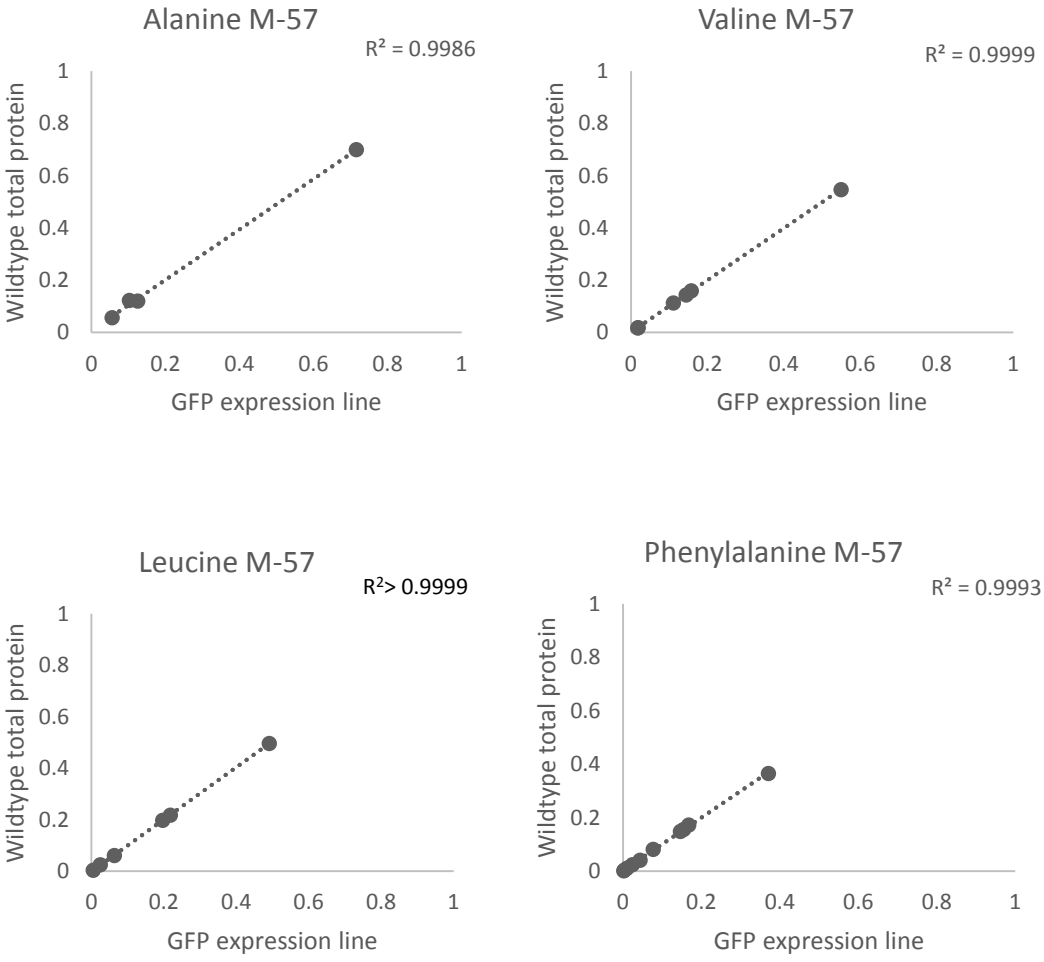


Figure 32: **Comparison of total protein mass isotopomer distributions from wildtype and GFP-expressing seedlings.** Total ion counts measured from amino acid mass peaks derived from the total protein fractions from wildtype seedlings and seedlings with constitutive GFP expression grown in 20% [$^{13}C_6$]-glucose

in hydroponic cultures were processed into mass isotopomer distributions to obtain average data on each mass peak corresponding to each fragment. These data were compared between the two total protein fractions by fitting a line through the data and calculating a coefficient of determination.

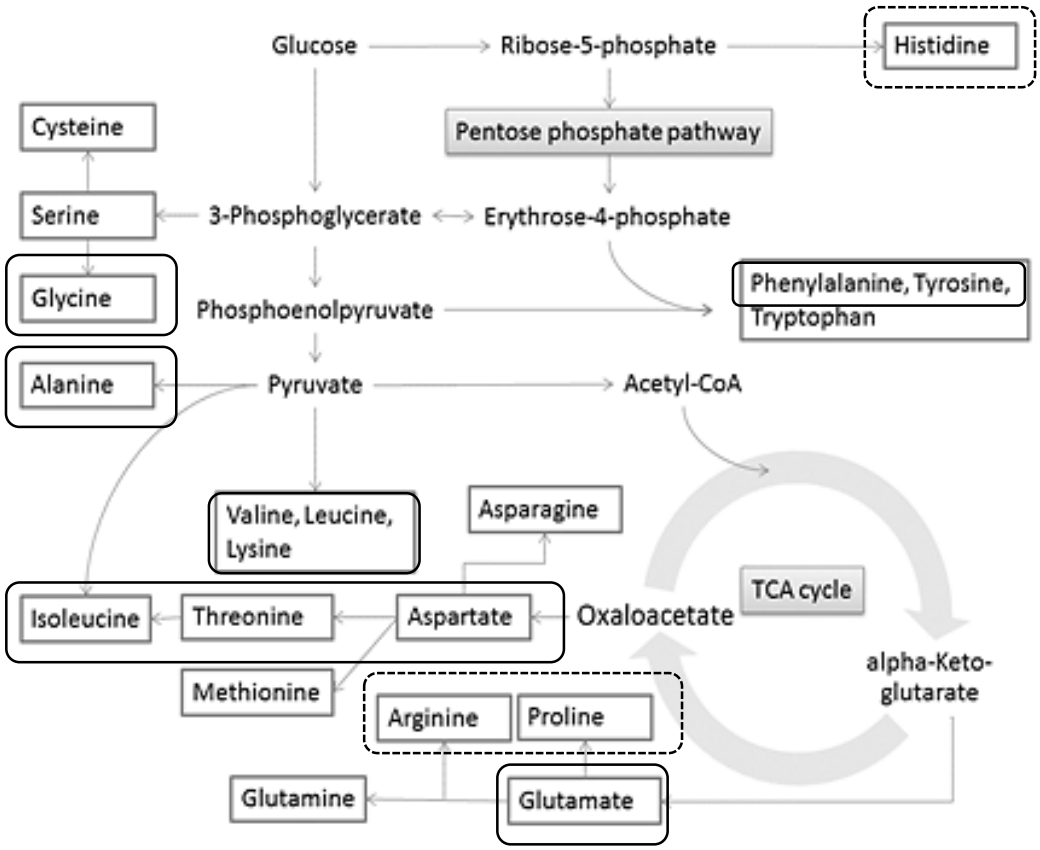


Figure 33: **Central carbon metabolism with the minimum set of fragments that could be reliably detected from immunopurified hydrolysed GFP shown.** The amino acids where consistent measurements could be obtained have been circled and the cut line indicates amino acids where data was measured for some fragments or different fragments were measured from the total protein and GFP samples. The data covers the major branches, if different fragments are used for detection of mass isotopomer distributions for histidine. Mass isotopomer distributions from arginine and proline were found to have lower coefficients of determination in the analysis.

5.3 Discussion

5.3.1 Amino acids derived from plant GFP reflect the mass istopomer distributions and fractional abundances of ^{13}C normally obtained from analysis of total protein

The mass isotopomer data detected from amino acids derived from purified GFP was for the vast majority of amino acid fragments in the analysis equivalent to the total protein fraction. This was true at the level of mass isotopomer distributions and calculated average fractional abundances of ^{13}C. Out of these, the mass isotopomer data represents the data used in metabolic models and the average fractional abundances provide a convenient assessment of the extent to which the analysed fragments have been derived from the labelled substrate.

However, for some amino acid fragments, the correspondence in mass isotopomer distributions between total protein from wildtype and GFP-expressing lines was less precise some amino acid fragments – this was reflected in lower coefficients of determination in comparisons between the data series. This discrepancy probably arises from difficulties in reproducibly measuring the data for each mass peak and quantifying the isotopic labelling, consequently causing variation in the fractional abundances for amino acids such as arginine and proline, and in the coefficient of determination for phenylalanine. While the variation in the measurement data is relatively small, variation in the data for heavier mass peaks is magnified through the process of calculating average ^{13}C percentages due to values being multiplied by the number of substituted carbons. This creates an additional source of variation with large fragments when comparing average ^{13}C enrichment. Both types of variation are more prominent in data obtained from GFP due to the lower amount of amino acids in samples, as shown by the larger standard deviations obtained for the data on average fractional abundances of ^{13}C. Irrespective of the precise explanation for the fluctuations in calculated average 13C enrichment, for the vast majority of the fragments, this variation does not significantly affect the ability to report on the isotopomer distributions in the amino acid pool which provide a more precise measure of the redistribution of label by the metabolic network

Variation in measurements is a factor included in models used for metabolic flux analysis. Data on fragments showing higher variation can therefore be used as part of the model, but its usefulness decreases as the variation in the data increases. An additional benefit of using amino acid data for models is also that the some of the measurements may be redundant in case several amino acids or fragments provide data on the same precursor pool. It is common not to detect or use all amino acid fragments in flux analysis experiments, arginine being one of the amino acids that is often difficult to detect reliably and therefore frequently not used in analysis. Thus, for example,, the only published paper on validation of GFP as a reporter protein that presents amino acid data does not mention arginine or proline (Shaikh et al., 2008). Similarly, in analysis of labelling in RuBisCO subunits, Allen and colleagues were unable to measure ^{13}C enrichment from methionine, lysine, arginine, histidine and tyrosine due to low levels (Allen et al., 2012).

The data reported by Shaikh and colleagues for the GFP expressed in E. coli consists of isotopomer distribution values for ten amino acids (alanine, glycine, valine, leucine, isoleucine, serine, phenylalanine, aspartate, glutamate and histidine) (Shaikh et al., 2008). The group used a single-labelled substrate and therefore detect mainly M0 and M+1 mass peaks. While this type of dataset could be enough to provide enough labelling information depending on the experiment, use of [$^{13}C_6$]-glucose is likely to generate a far greater range of mass isotopomers, which can provide a more informative source of data. The coverage of the network by the fragments used in this part of the analysis is shown in Figure 33 .

5.3.2 Arabidopsis seedlings grown in hydroponic cultures reach an isotopic steady state with minimal contribution from CO_2 through photosynthesis

In the experiments conducted in this chapter, the average fractional abundance of ^{13}C in the vast majority of analysed fragments suggest that the amino acids in protein extracted from Arabidopsis seedlings had reached the 20% ^{13}C enrichment that equivalent to the isotopic labelling of the substrate. Curiously, this was the case with protein extracted from not only roots but also leaves, suggesting that photosynthesis was negligible under the culturing conditions used in this study.

It is unclear why photosynthetic carbon assimilation did not make a measureable contribution to leaf metabolism in the current culture system. The seedlings developed recognisable leaves that remained green up until harvest at 14 days and showed no obvious morphological defects. It is possible that the light intensity used in the standard culture conditions was sufficient to trigger normal photomorphogenic development but was too low to drive the light reactions of photosynthesis at a significant rate. Both weak and strong light conditions have been shown to cause photoinhibition when Arabidopsis seedlings are cultured *in vitro* (Eckstein et al., 2011). In addition, sugars also have an effect on expression of genes encoding RuBisCo and other enzymes of the Calvin cycle (Paul and Pellny, 2003)

In summary, the seedlings reach isotopic steady state within the 14 days of culturing, which is observed as approximately 20% fractional abundance of ^{13}C. Photosynthesis is unlikely to confound analysis of label redistribution in roots in this system. However, studies aiming to target normal leaf metabolism will be compromised by the apparent lack of photosynthesis under the culture conditions used. However, studies aiming to target normal leaf metabolism will be compromised by the apparent lack of photosynthesis under the culture conditions used.

5.3.3 GFP can be used as a reporter protein in the roots of Arabidopsis seedlings grown in hydroponic cultures

Two main results of this chapter demonstrate the suitability of GFP to be used as a reporter protein in the Arabidopsis root under the chosen experimental conditions. First, the comparability of the mass isotopomer data obtained from total protein and GFP establish that the latter may be used as a reliable source of measurements in cell-type-specific MFA in this system. Secondly, there was no indication of any effect from the GFP over expression on the isotopomer data obtained from seedlings with constitutive expression of GFP. In terms of the experimental system, it was found the contribution of CO_2 assimilation from photosynthesis to the isotopic labelling in the system was minimal and the average fractional abundances of ^{13}C indicate the amino acid pools used to synthesise protein had reached a ^{13}C enrichment of approximately 20%. Finally, showing that

measurements from amino acid fragments derived from a diverse range of network intermediates can be obtained from the GFP reporter protein demonstrates the suitability of the approach for steady state stable isotope MFA.

5.3.4 Use of isotopomer data obtained from a reporter protein for modelling metabolic activity

A successful validation is a prerequisite for using reporter protein data for modelling cell type specific metabolic activity. In addition, the data obtained from the amino acid fragments need to cover enough of the metabolic network to reliably report on the different branches of central carbon metabolism. Ideally several fragments that can report on a single precursor pool should be detected and measured reliably to satisfy the requirement to over determine the model used for the computational fitting of data in MFA experiments. When available data are limited, other approaches that have been used include flux ratio analysis, that has been demonstrated to provide information on the metabolic activity of specific cell types through the reporter protein strategy in micro-organisms (Rühl et al., 2011).

A recent development in work with micro-organisms is the determination of flux distributions from measurements of the isotopic profiles of free amino acids, which have been shown to be similar to those of amino acids derived from protein (Okahashi et al., 2014). The advantage of this approach is that the soluble pools reach isotopic steady state up to ten time faster than the proteins. The authors conclude the necessary minimum measurements required were data on glutamate, aspartate, alanine and phenylalanine. With this dataset, they were able to deduce flux distributions containing estimates for fluxes through the precursors of these amino acids: pyruvate, oxaloacetate, alpha-ketoglutarate, phosphoenolpyruvate and erythrose-4-phosphate. The reporter protein strategy described in this chapter would allow detection of datasets similar to those used by Okahashi and group, provided the levels of GFP purified from the cells of interest is high enough. Here data on all of these amino acids could be measured from GFP derived from the line with constitutive GFP expression.

Inherently the reporter protein strategy is limited to measuring data that is available through the amino acids contained in proteins. However, conventional MFA uses measurements from other

metabolites such as sugars, polysaccharides and other biomass components. In the vast majority of cases, obtaining data on these from specific cell types is just as difficult as obtaining data on protein from specific cell types without the use of a reporter or alternative techniques such as microdissection or flow cytometry, which are time consuming and may affect the measurements. Chapter 6 describes work on one cell type where such measurements are possible: root hairs of *Arabidopsis thaliana* seedlings. Building models based on the data obtained from such systems can ultimately aid in deciding the best modelling approaches to employ when data are limited to isotopomer measurements obtained from protein hydrolysates.

While not currently available, novel methods for spatially-resolved measurements of either the levels or stable isotope labelling of biomass components may arise through techniques such as atmospheric infrared MALDI, which is already capable of detecting over 50 small metabolites from the surface of plant tissues (Li et al., 2008b). Detection of stable isotope labelling from biological tissue by these types of techniques presents a complex problem to be solved, but encouraging development is being made with applications such as detection of growth of antibiotic resistant micro-organisms in the clinical setting with MALDI-TOF MS using stable isotope labelling (Sparbier et al., 2013). These novel techniques may ultimately be able to provide more data on specific cell types or identify how large the differences between them are. However, currently such techniques are at early stages of development and it is unclear when, and to what extent, they will be able to provide the type of data needed for flux analysis. In the meantime, the cell-specific reporter as described here remains the only practicable strategy.

6 Root hair cells as a model system for investigating cell type specific metabolism

The previous chapters have focused on validation of the reporter protein strategy for use in *Arabidopsis thaliana* roots and described the requirements and tools to be used in such analyses through *in silico* models and experimental validation. This chapter describes the work towards developing a model system for cell type specific metabolism using isolated root hairs.

6.1 Introduction

Techniques that aim to study single cells or specific cell types are often based on isolation of cells from the tissue. Cells that can be separated without effect on the biology under investigation can be analysed using standard techniques in a way similar to cells grown in culture. In plants hair cells are one such cell type that can be easily isolated, as they protrude from the tissue surface. Hair cells are specialised epidermal cells that grow out of the leaf, stem or root as thin extensions. In the roots this increases the area that is in contact with the soil.

Isolation of root hairs from the roots can be achieved by stirring in liquid nitrogen (Qiao and Libault, 2013) and therefore does not add any experimental steps that might affect the cell metabolism detected through common techniques used for metabolic flux analysis. Developing a model system based on root hairs could benefit investigations of cell type specific metabolism in several ways. Firstly, additional metabolic information can be obtained from isolated hair cells compared to a system reliant entirely on analysis of purified reporter protein. This additional information may allow the evaluation of how well protein alone reports on metabolism in the cells compared to the variety of metabolite data used to derive a standard flux map. Secondly, testing different modelling approaches in the root hair system could guide the development of models for cell type specific work and help in further validation of strategies. Finally, root hairs are a biologically interesting cell type with an important role in transport of water and nutrients (Gilroy and Jones, 2000), and more broadly have been used as a model system to study growth and development (Foreman and Dolan, 2001; Schiefelbein, 2000). The result of this focus is a steadily increasing understanding of the complex regulatory networks in root hairs and the discovery of different phenotypes dependent on expression of specific genes (Becker et al., 2014; Foreman and Dolan, 2001; Lan et al., 2013). Root hair cells

have been suggested as the ideal model for systems level approaches to study plant cells (Libault et al., 2010; Qiao and Libault, 2013) and metabolic flux analysis can provide an important contribution to this type of systems analysis (Sweetlove et al., 2008).

Investigation of intracellular fluxes in root hairs is additionally supported by recent discoveries at other levels of the metabolic network. The focus on root hair cells has resulted in datasets for gene and protein expression using fluorescence activated cell sorting (Lan et al., 2013). The analysis of these datasets showed an interesting difference between root hair cells and the mixed cell fraction: the abundance of transcripts for several of the genes encoding enzymes from the primary metabolism are enriched in root hairs. These include the cytosolic phosphoenolpyruvate carboxykinase 1 (catalysing the reaction of oxaloacetate into phosphoenolpyruvate), the $NADP^+$-dependent malic enzyme (catalysing the reaction of malate into pyruvate) and alcohol dehydrogenase 1. Levels of peptides derived from the proteome were determined with LC-MS/MS analysis and revealed increases in the levels of proteins linked to the genes enriched in root hairs. The authors suggest the results indicate several genes linked to activation of metabolic pathways required for energy production and cell wall synthesis were highly enriched in root hair cells. Metabolic flux analysis could provide validation of these proposals by direct analysis of the flux phenotype of root hairs.

The work on growth and development has also led to characterisation of several genes that are involved in root hair formation and growth, including the RSL4 transcription factor (ROOT HAIR DEFECTIVE 6-LIKE 4, (Yi et al., 2010)). RSL4 over expression leads to longer root hairs, while lack of expression results in very short hairs. Under normal conditions, RSL4 regulates root hair growth in response to external signals such as low phosphate availability and auxin and therefore provides a link between environmental signals and post-mitotic growth (Yi et al., 2010). Growth of longer root hairs is known to affect the uptake of inorganic phosphate, and root hair length depends on the concentrations of phosphate available in the environment surrounding the root (Niu et al., 2013). Longer root hairs may well provide a physiological advantage by increasing the area of the roots, but the effect of over expression of RSL4 and the longer root hair phenotype on metabolism in root hair cells has not been established.

Root hairs provide a unique system in which cell type specific flux analysis is amenable to direct interrogation by physical isolation of the specific cell type. Flux maps generated from data obtained from isolated cells could then be compared with those produced from data gained from analysis of GFP expressed exclusively in root hairs to provide a direct test of the extent to which the reporter protein strategy is able to provide an accurate flux map of the targeted cell type. The major aim of this chapter was to investigate the ^{13}C labelling in amino acids derived from isolated root hairs in order to assess the suitability of this system for use as a model system for cell type specific metabolism. This model system could additionally link the work on cell type specific metabolic fluxes with other approaches aiming to elucidate metabolic activity of root hairs at the level of different cell types

The specific objectives were: 1) to compare the fractional abundances of ^{13}C from amino acids measured from the whole root total protein fraction and the total protein fraction from root hairs separated from Arabidopsis seedlings; 2) to compare the detected fractional abundances of ^{13}C in metabolites from root hairs from wildtype seedlings and metabolites from root hairs isolated from a line with over expression of the RSL4 transcription factor; and 3) to compare the levels of metabolites in exudates between the over expression line and wildtype seedlings.

Focusing on molecules relevant to central carbon metabolism, the chapter shows quantification of metabolite levels from exudates and ^{13}C enrichment in metabolites for a specific cell type (root hairs) isolated from lines with different phenotypes (wildtype and RSL4 over expression). A comparison of these data to wildtype whole root total protein fraction is presented in order to evaluate the feasibility of using root hair cells as a model system for cell type specific MFA and wider work on cell type specific metabolism. The work forms the basis for applying metabolic flux analysis in this system and further development of approaches for cell type specific MFA.

6.2 Results

6.2.1 Stable isotope labelling in root hairs is distinct from the labelling in whole root for a group of amino acids derived from the total protein fraction

To quantify the isotope enrichment in the root hair total protein fraction, wildtype Columbia seedlings were grown in the low phosphate Johnson's medium with 20% [$^{13}C_6$]glucose and 80% natural D-glucose. The root hairs from 14-day-old seedlings from a single culture of 30-50 seedlings were separated after harvesting in liquid nitrogen. The protein from the root hairs and the whole root total protein fraction were hydrolysed separately, derivatised and analysed with GC-MS. The levels of amino acids detected from the root hair sample were at similar levels to those normally detected from total protein samples, suggesting enough material had been obtained during the isolation step (Figure 34). The average fractional abundance of ^{13}C in each amino acid fragment was calculated and amino acid fragments detected in only one sample type or producing inconsistent or unreliable mass isotopomer distributions were discarded from subsequent analysis.

There was no significant difference in the extent of labelling of amino acids derived from root protein of seedlings grown in MS medium and Johnson's medium (Figure 36). Both protein fractions yielded the expected value of 20%. Similar to this, fractional ^{13}C abundance of all amino acids in protein extracted from whole roots were close to that of the supplied substrate, but the values for several amino acid fragments in protein derived from root hairs were significantly lower than the expected value of 20% (Figure 35). The reduction detected from root hair total protein is statistically significant ($p<0.05$) for all aspartate and glutamate fragments and some of the phenylalanine, serine and methionine fragments when compared with the whole root total protein fraction (Figure 36, statistical analysis of the data including correction for false discovery rate is detailed in Appendix 21)

One of the advantages of using root hair cells for the analysis is the ability to detect the isotope labelling from a variety of metabolites such as sugars and organic acids, which are included as part of the dataset for standard MFA experiments. If this data can be measured from isolated cells, a flux map produced through standard techniques can be compared to a cell type specific flux map based on protein labelling alone. This type of direct comparison between the reporter protein strategy and conventional MFA could help guide the modelling approaches for cell type specific flux analysis. The ability to physically isolate this specific cell type using a relatively simple and rapid technique provides an unparalleled opportunity to compare flux maps generated using contrasting strategies. Experiments on soluble metabolites can additionally confirm that the labelling pattern in amino acids in protein in a specific cell type accurately reflect the labelling patterns in the metabolic intermediates in those cells. The fraction of methanol soluble metabolites was separately analysed with GC-MS and the metabolites present were identified against the National Institute of Standards and Technology (NIST) metabolite database before processing the data. The methanol soluble fraction extracted from root hairs contained sugars, organic acids and amino acids that all contained approximately 20% fractional abundance of ^{13}C (Figure 37). The free glutamate and serine that were detected from the soluble fraction did not show the lower ^{13}C enrichment detected from amino acids derived from protein

6.2.2 Influence of RSL4 over-expression on the fractional abundance of ^{13}C detected in root hairs in seedlings grown in low phosphate

To investigate the effect of RSL4 over expression on fractional abundances of ^{13}C detected from root hair cells, mass isotopomer distributions were obtained from the amino acids derived from separated root hairs of the over expression line and Columbia wildtype seedlings and processed into fractional abundances of ^{13}C. All seedlings used in the experiments were grown in the low phosphate Johnson's medium where the long root hair phenotype is observed with 20% [$^{13}C_6$]-glucose as the substrate.

The fractional abundances of ^{13}C in amino acid fragments derived from protein extracted from the root hairs in the RSL4 line did not show any differences to wildtype root hairs (Figure 38). Similar

results were obtained when the protein extracted from isolated root hairs from the RSL4 line were compared against the wildtype total protein fraction (Figure 39). Statistical analysis did not confirm any of the variation to be significant between the two datasets, with $p<0.05$ used as the significance cut off.

To further understand the labelling patterns detected from the root hair samples, mass isotopomer distributions were analysed for 13 metabolites from the methanol soluble fraction, including amino acids, sugars and organic acids. The average ^{13}C percentage in these metabolites indicated most of them contained approximately 20% ^{13}C (Figure 40). A comparison of this data from the over expression line with the fractional abundances of ^{13}C detected from the methanol soluble fraction derived from wildtype root hairs indicated no significant differences between the two sample types.

6.2.3 Comparison of exudate composition between wildtype and RSL4 over expression seedlings

The long root hair phenotype of the seedlings over expressing RSL4 has not been characterised in terms of the effect of the over expression on exudate composition. Differences in secreted metabolites could provide preliminary data on whether the long root hair phenotype is linked to any functional changes in these cells. Any observable change in fluxes, for example, could be exploited in validation of cell type specific MFA in the future, in addition to being of interest in terms of the root hair biology.

The main limitation for studying metabolites in the exudates is that the medium the plants are grown in is rich in nutrients and sugar. Because of this, several compounds are present at much higher levels than the metabolites of interest secreted by the plants. Running samples with concentrations of these low abundance metabolites at high enough levels to quantify them will overload the mass detector with the components of the growth medium during GC-MS analysis. This problem can be overcome by growing plants in growth medium and then transferring them to water for a set period of time to collect exudates (Badri et al., 2013). By adapting this method to liquid cultures as described in Chapter 2, it was possible to collect water samples that did not contain the nutrients that were

interfering with detection from the growth medium and collect information on the composition of exudates.

Exudate composition was analysed from the data collected from freeze dried samples by identifying peaks in the spectra using the NIST metabolite database and quantifying the peaks against an internal standard (ribitol). To take into account how many seedlings had germinated out of the 30-50 seeds, the fresh weight was recorded and used to calculate the amount of each metabolite secreted into the medium per gram fresh weight. Due to differences in seedling size in the cultures and the error in counting individual seedlings from liquid cultures, fresh weight was considered a more accurate basis on which to express the production of root exudates. The quantities of metabolites in exudates were compared between the RSL4 over expression line and Columbia wildtype seedlings (Figure 41). After correction for false discovery rate, metabolite levels were found to be indistinguishable in the two types of exudate samples. Therefore comparison of the exudate levels did not reveal any evidence of differences resulting from the over expression of RSL4 and the longer root hair phenotype.

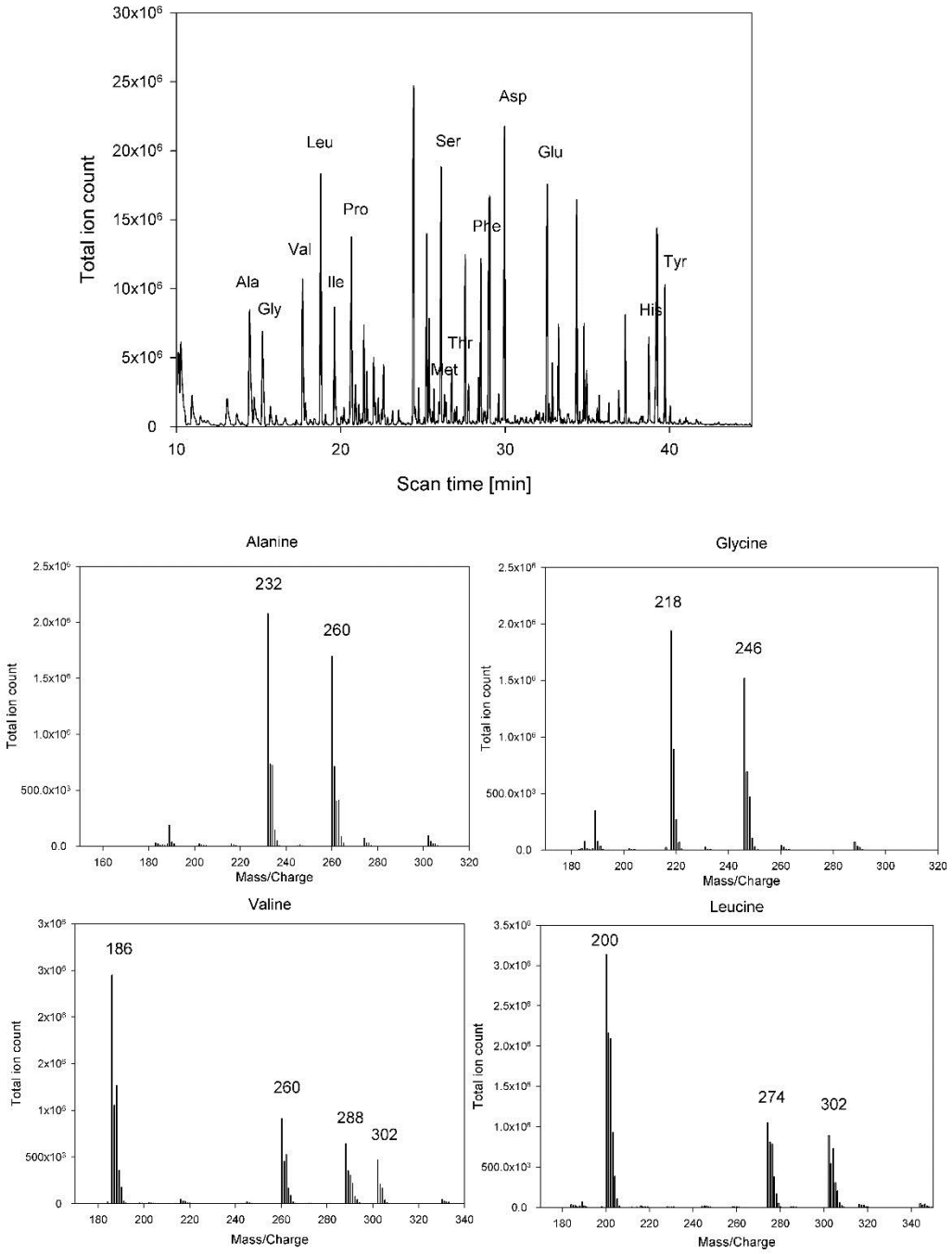


Figure 34: **Chromatographic detection of amino acids from total protein derived from separated root hair cells.** Protein extracted from root hairs grown in hydroponic cultures and separated in liquid nitrogen was hydrolysed, derivatised and analysed by GC-MS using conventional conditions. Representative mass spectra for amino acids from the root hairs grown in 20% [$^{13}C_6$]-glucose are shown.

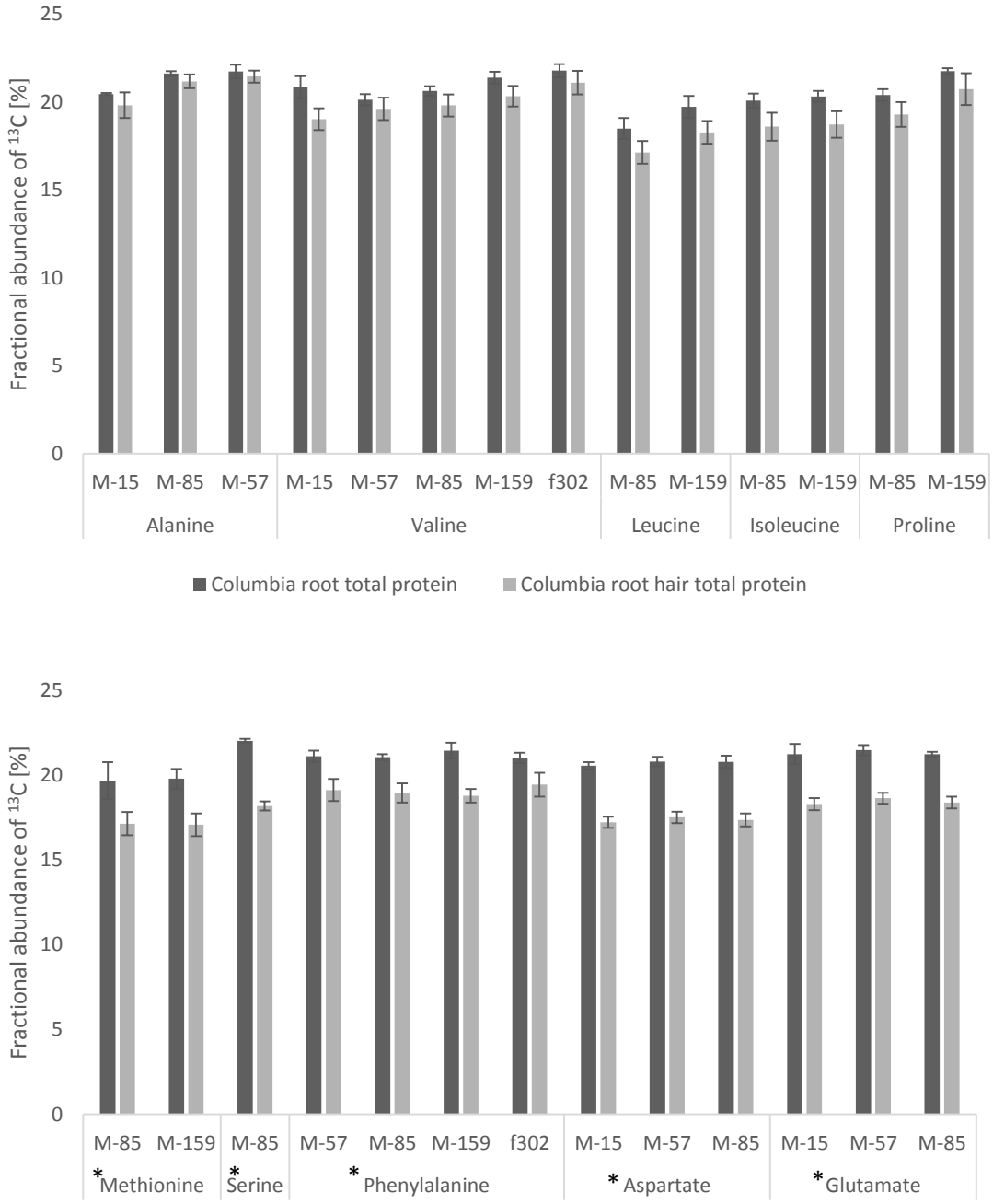


Figure 35: **Comparison of average fractional abundances of ^{13}C detected from amino acid fragments in isolated root hairs and the whole root total protein fraction from Arabidopsis seedlings grown in 20% [$^{13}C_6$]-glucose in Johnson's medium (low phosphate).** Data are displayed as the mean and standard deviation of fractional abundances measured from amino acids fragments isolated from root hairs and the whole root total protein fraction. Data from a minimum of four samples from biological replicates were used for each amino acid fragment. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets. Amino acids where measurements from the fragments were different between the two dataset ($p<0.05$) are indicated with *

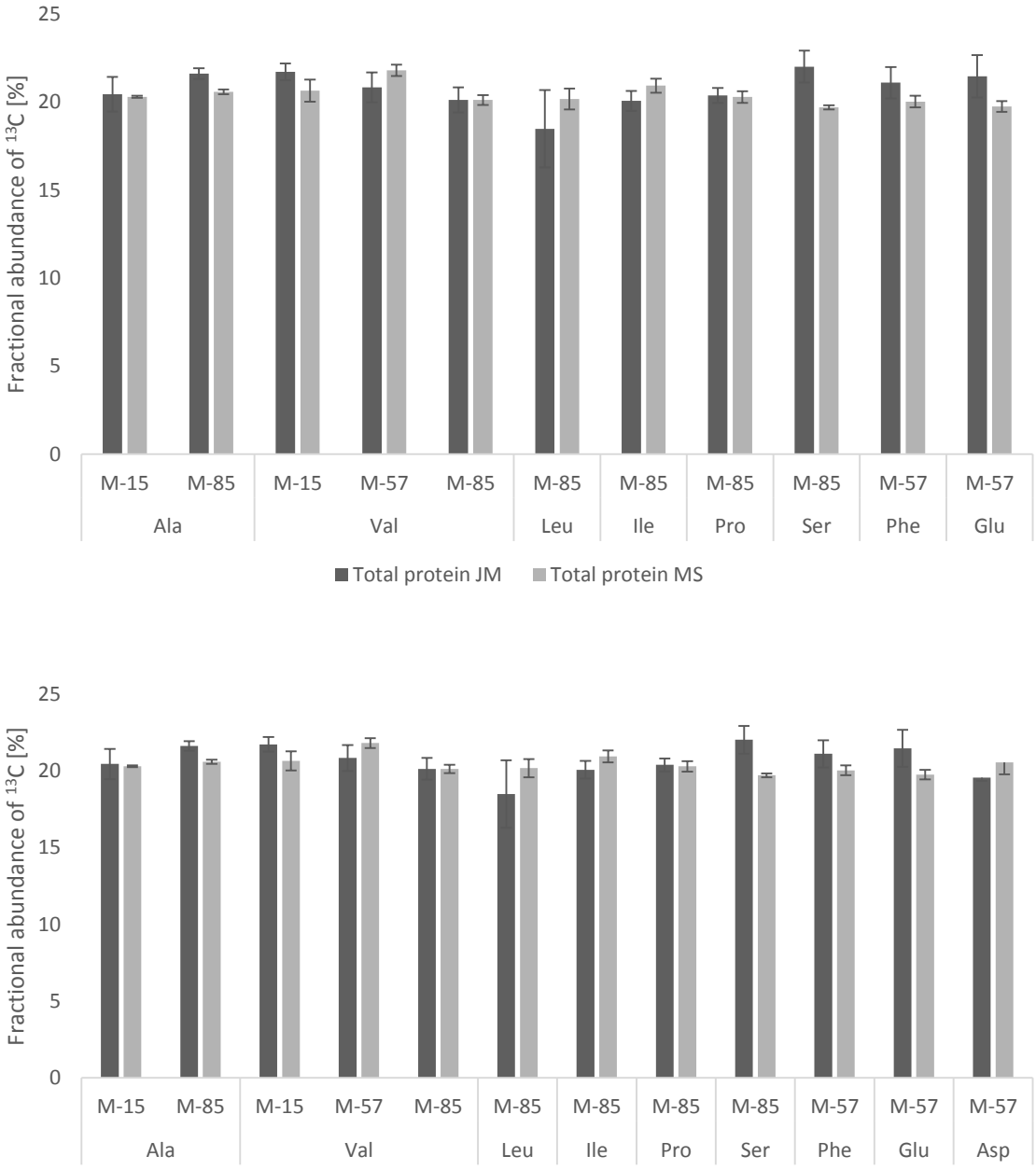


Figure 36: **Comparison of average fractional abundance of ^{13}C in whole root total protein fractions from seedlings grown in Johnson's medium (low phosphate) and MS medium.** Columbia wildtype seedlings were grown in hydroponic cultures of 30-50 seedlings in either Johnson's medium or MS medium providing 20% [$^{13}C_6$]-glucose with D-glucose as a substrate. Data are displayed as the mean and standard deviation of fractional abundances measured from amino acids fragments derived from the root total protein fraction. Data from a minimum of six samples from biological replicates were used for each amino acid fragment. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

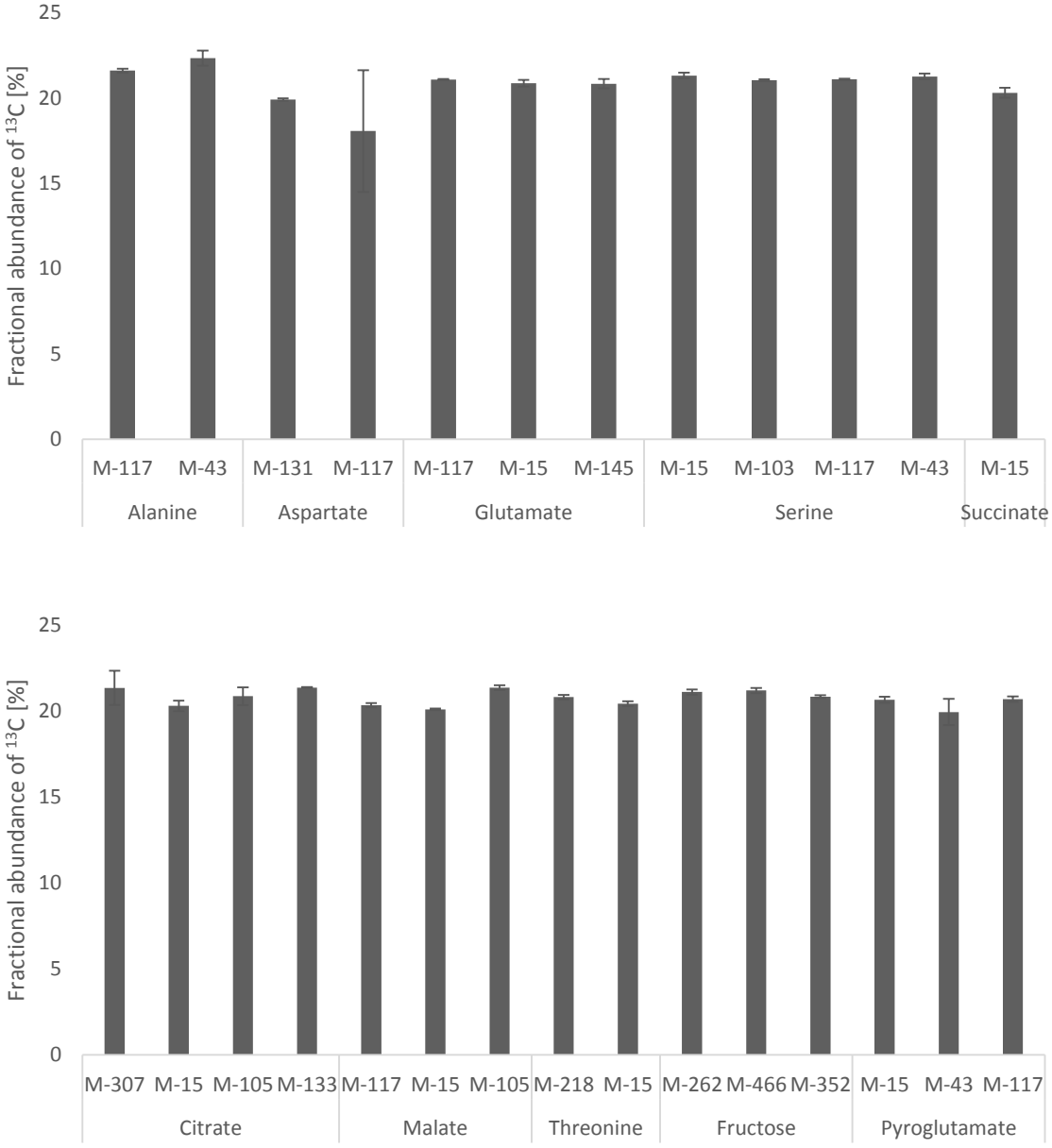


Figure 37: **Average fractional abundances of ^{13}C in methanol soluble metabolites derived from the total protein fraction in isolated root hairs.** Columbia wildtype seedlings were grown in hydroponic cultures of 30-50 seedlings in Johnson's medium with 20% [$^{13}C_6$]-glucose provided as substrate. Root hairs were isolated in liquid nitrogen and the soluble fraction was obtained from methanol extraction of all of the biological material. Data are displayed as the mean and standard deviation of fractional abundances measured from metabolites. A minimum of four samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

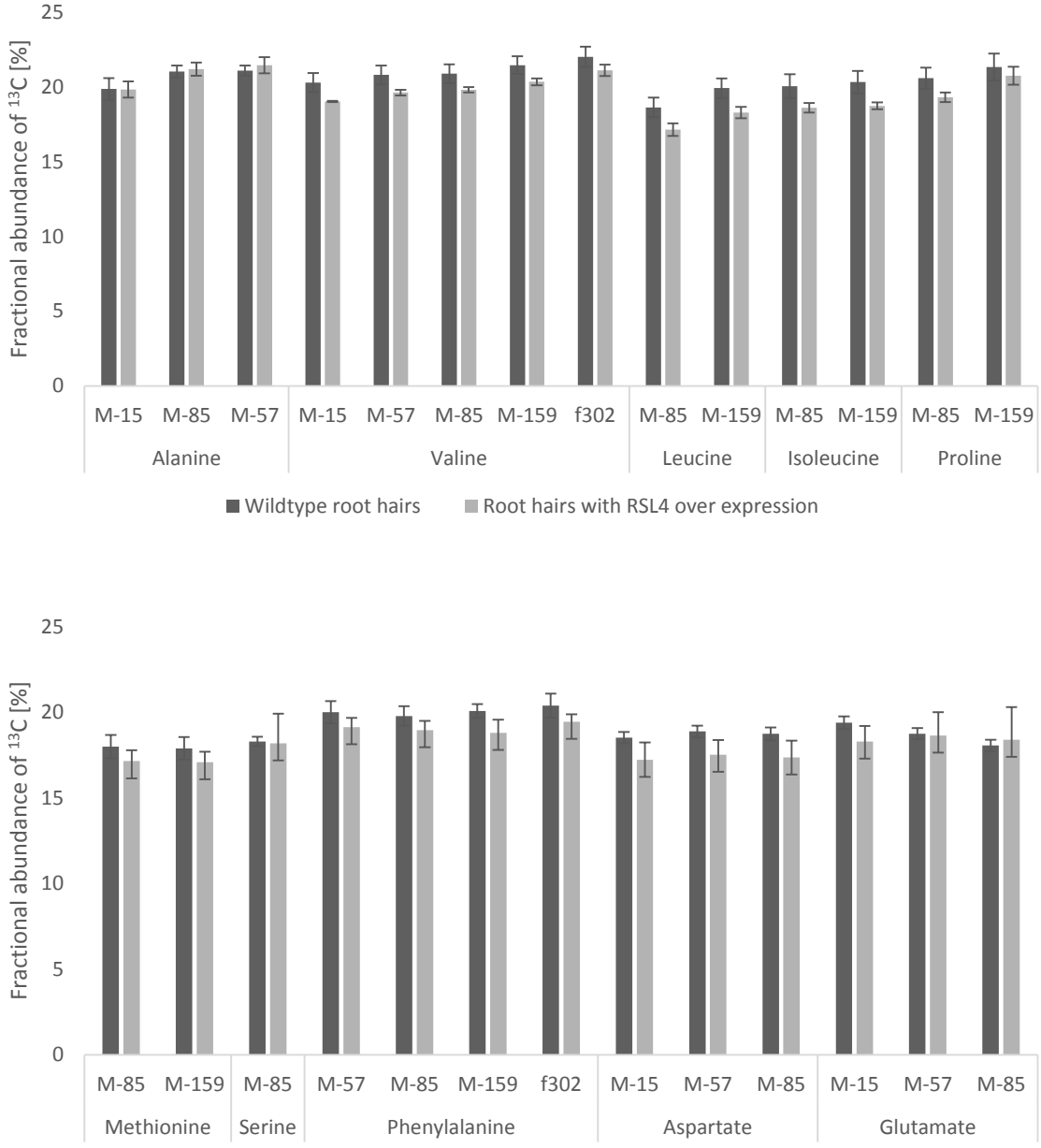


Figure 38: **Comparison of average ^{13}C fractional abundance of amino acids derived from total protein fractions of isolated root hairs from wildtype seedlings and from seedlings with RSL4 over expression.** Arabidopsis seedlings were grown in 20% [$^{13}C_6$]-glucose in Johnson's medium and root hairs were isolated in liquid nitrogen. Protein samples were hydrolysed, derivatised and analysed with GC-MS using standard conditions. Data are displayed as the mean and standard deviation of fractional abundances measured from amino acid fragments. A minimum of six samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

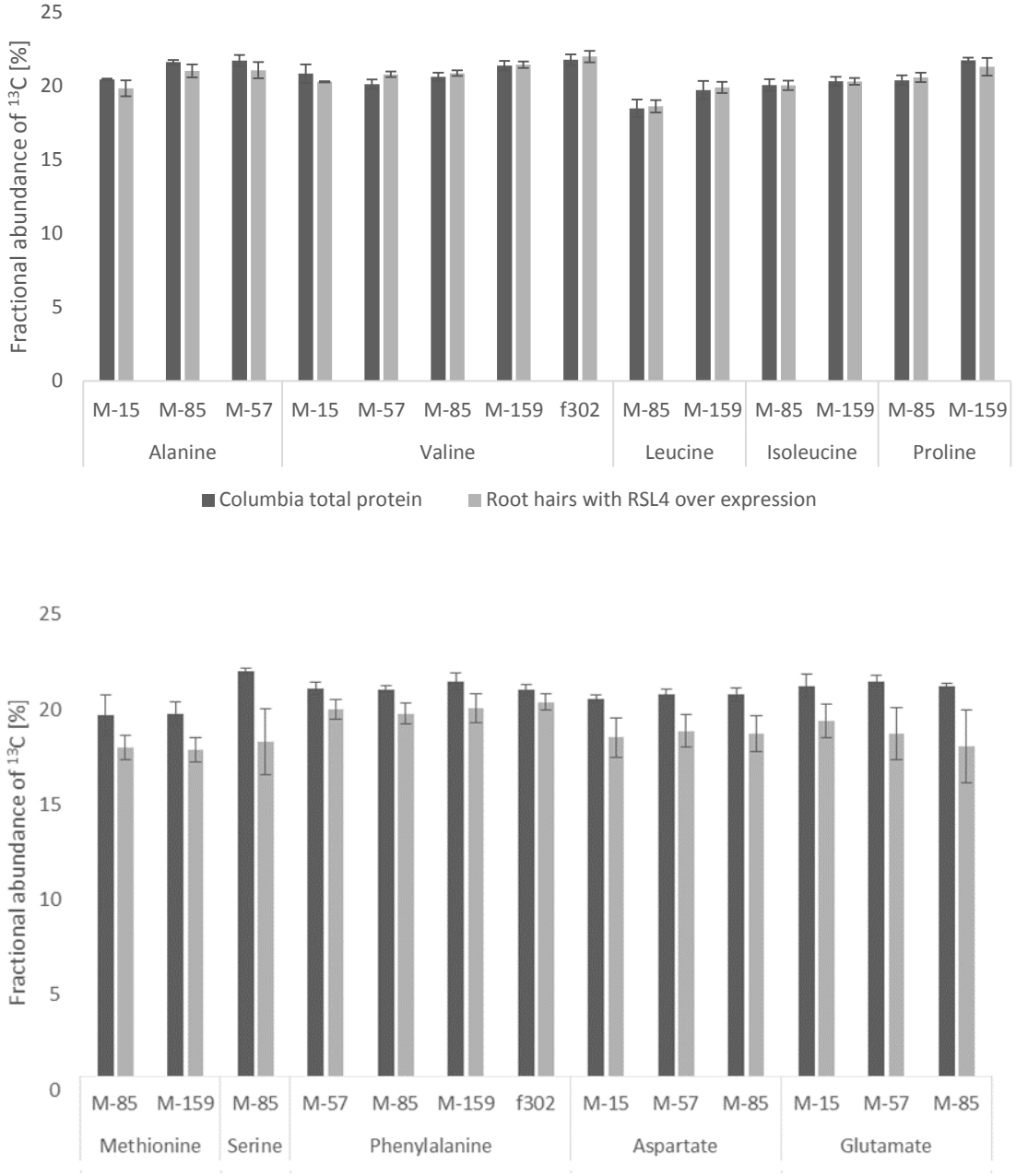


Figure 39: **Comparison of average fractional abundances of ^{13}C detected from amino acid fragments in whole root total protein fractions from wildtype seedlings and seedlings with RSL4 over expression.** Arabidopsis seedlings were grown in 20% [$^{13}C_6$]-glucose in Johnson's medium. Protein was hydrolysed, derivatised and analysed with GC-MS using conventional conditions. Data are displayed as the mean and standard deviation of fractional abundances measured from amino acid fragments. A minimum of six samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

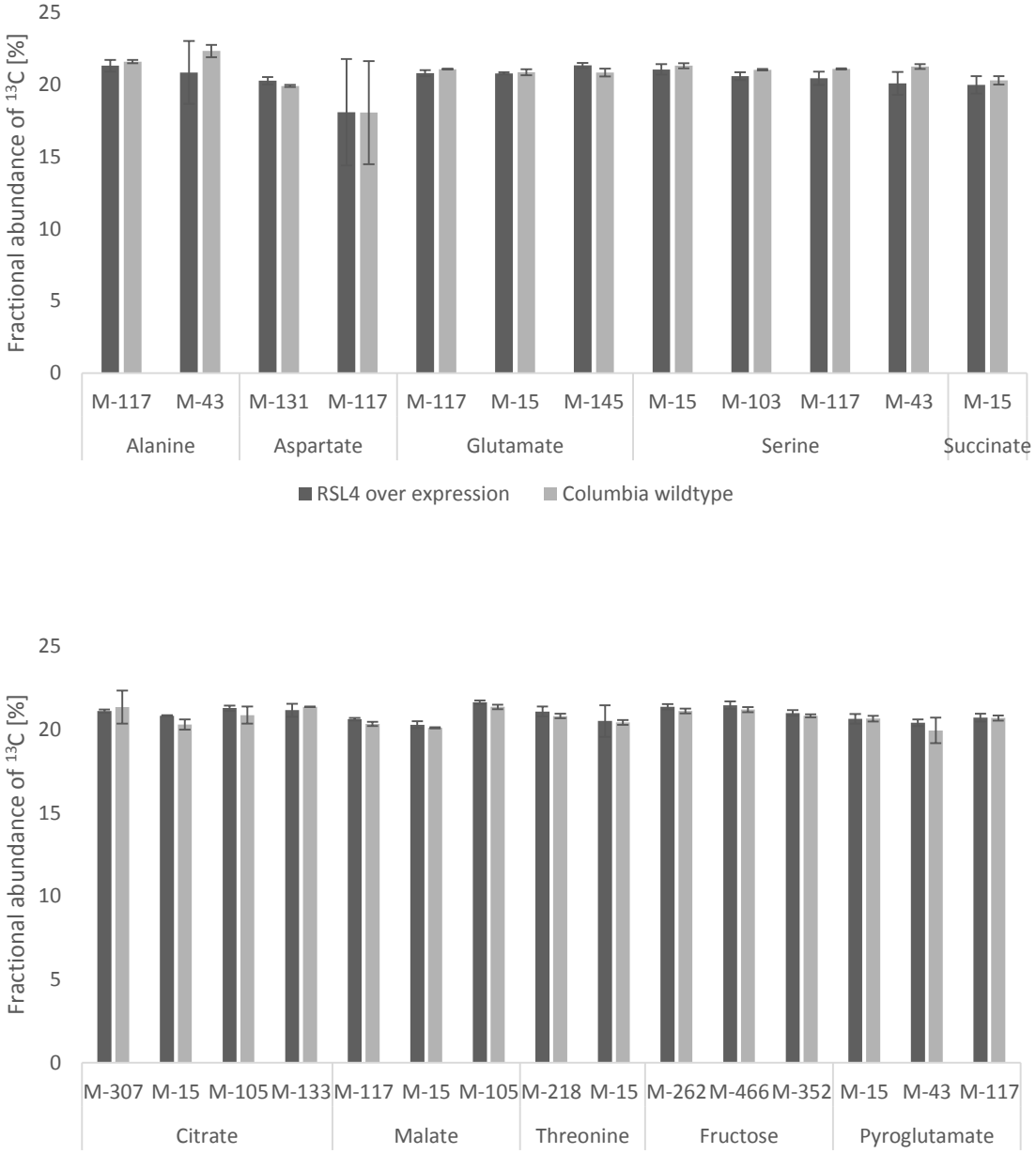


Figure 40: **Comparison of average fractional abundances of ^{13}C detected from methanol soluble metabolites in root hairs isolated from wildtype seedlings and seedlings with RSL4 over expression.** Arabidopsis seedlings were grown in 20% [$^{13}C_6$]-glucose in Johnson's medium. Seedlings were grown in hydroponic cultures and root hairs were isolated in liquid nitrogen. The material was ground in liquid nitrogen and the methanol soluble fraction was obtained from all of the biological material. Data are displayed as the mean and standard deviation of fractional abundances measured from metabolites. A minimum of four samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

6.3 Discussion

6.3.1 ^{13}C enrichment suggests root hair proteins are not labelled to isotopic steady state

For labelling experiments that rely on uniformly labelled substrates such as [$^{13}C_6$]-glucose as the sole carbon source, the end products that are derived from the substrate are expected to reflect the fractional abundance of label in the substrate. Deviation from the expected average fractional enrichment could indicate problems with data analysis or analytical detection of the mass isotopomers. Additionally, factional abundances lower than predicted from that of the substrate may result from the contribution of pools of metabolites that are present at the beginning of the labelling experiment.

In the total protein fractions obtained from isolated root hairs, the average fractional abundances of ^{13}C for methionine, serine, phenylalanine, aspartate and glutamate fragments were lower than predicted from the 20% [$^{13}C_6$]-glucose provided in the culture medium. It is unlikely that differences in growth conditions, sample preparation of GC-MS analysis contributed significantly to this variation since biological replicates of the different samples were grown and prepared at the same time and run in random order; moreover, comparisons of whole root protein and root hair protein were performed on samples derived from the same plants. Equivalent amounts of protein were analysed from the two sources to ensure that the levels of amino acids detected from the two sample types were very similar and varied within the same range (Figure 34). Baseline correction and mass correction were applied to the two sets of data at the same time and sample type was only taken into account at the final step when data were compared. Therefore, it seems unlikely the differences between sample types were caused by systematic differences in either data analysis or the analytical technique.

Metabolic flux analysis normally requires an isotopic steady state, where the level of isotopic labelling in the metabolite pools under investigation is constant. Biomass components, such as protein, take longer than low molecular weight intermediary metabolites to reach isotopic steady state. The pools of soluble metabolites in a typical plant cell usually reach steady state in a few hours,

(Troufflard et al., 2007; Williams et al., 2008). Contribution of lipids as a respiratory substrate during germination and early seedling growth will dilute the label in the pools of metabolic intermediates during ^{13}C labelling experiments and thereby increase the time needed for soluble amino acids (and the proteins derived from them) to achieve the isotopic abundance of the supplied exogenous substrate. In Arabidopsis seedlings existing metabolite pools present in seeds contain storage lipids, protein, organic acids and amino acids (Lonien and Schwender, 2009). These storages would typically be depleted within three days from the start of germination (To et al., 2002). In Arabidopsis cell cultures, cells reach isotopic steady state in the intact protein within five days (Williams et al., 2008). Storage lipid breakdown under different growth conditions can provide some indication of the different timescales that have been observed previously: Arabidopsis seedlings grown in cultures with high glucose concentrations (0.3M) in the medium retain 80% of the seed storage lipids 22 days from germination, while normally majority of the storage lipids would have been broken down by day 10 (To et al., 2002). The cells in Arabidopsis seedlings therefore do possess the flexibility that might make it possible for the root hair cells to take longer than the average cell in the Arabidopsis root to reach isotopic steady state and therefore contain lower isotope enrichment after 14 days in culture.

Approximately 20% fractional abundances were detected across amino acids derived from the combined whole root total protein fractions from seedlings in all of the experiments described in this thesis. This indicates that on average an isotopic steady state was reached within the 14 days of growth in 20% [$^{13}C_6$]-glucose. However, if all or some of the cell types synthesise their own amino acid pools, they may reach isotopic steady state at different time points and the labelling of the total protein fraction taken from a mixture might not be able to reflect this.

Even though there generally is a requirement for isotopic steady state in MFA experiments, the protein derived amino acids themselves do not need to be in an isotopic steady state, as long as there is a technique for assessing the labelling patterns of the precursors from which they are formed from the measurements made. If the amino acids derived from protein fail to reach an isotopic steady state, but the free soluble amino acids show higher labelling corresponding to isotopic steady state, it can

be assumed that the lower labelling results from an unlabelled protein pool that was present at the start of the experiment. This pool can be taken into account in modelling and its abundance can be determined by the fitting process (Kruger et al., 2012; Lonien and Schwender, 2009; Williams et al., 2008).

6.3.2 Labelling in amino acids derived from root hair protein and exudates from seedlings with long root hair phenotype is indistinguishable from wildtype seedlings

The exudates from Arabidopsis seedlings with long root hair phenotype caused by over expression of RSL4 were analysed to collect preliminary data on any potential changes in the metabolic activity or functions of the cells in comparison to wildtype. The data did not provide any evidence of differences between the line with RSL4 over expression and wildtype. This suggests that the increase in area of the root hairs might not have any obvious effect on the secretion of metabolites.

Previous studies of exudates have revealed that Arabidopsis seedlings with Arabidopsis seedlings with a no root hair phenotype have lower levels of metabolites in the exudates (Tanaka et al., 2014a) and especially the levels of secreted malate and citrate are decreased by 9% and 16%, respectively, compared to wildtype, when grown in low phosphate medium. Differences in citrate secretion and sugar metabolism have also been identified as important factors contributing to tolerance to P-deficiency in mutant plants of *Lupinus albus* (Li et al., 2008a). Based on the data collected here, simply increasing the area of the root by having longer root hairs might not have a corresponding effect of causing similar changes in the secreted metabolites. Interesting future work in this system could shed light on what, if any, effect the long root hair phenotype might have on the metabolic fluxes in these cells.

When root hairs from the seedlings with a long root hair phenotype grown in Johnson's medium with 20% [$^{13}C_6$]-glucose with over expression of RSL4 were isolated and the average fractional abundance of ^{13}C measured from amino acids derived from the total protein fraction of the isolated root hairs, these data showed no difference to the total protein fraction of wildtype seedlings. Both protein fractions and the metabolite soluble metabolite fraction reached the expected approximately

20% average ^{13}C enrichment. This is promising for the use of traditional MFA approaches to investigate the root hair phenotype further.

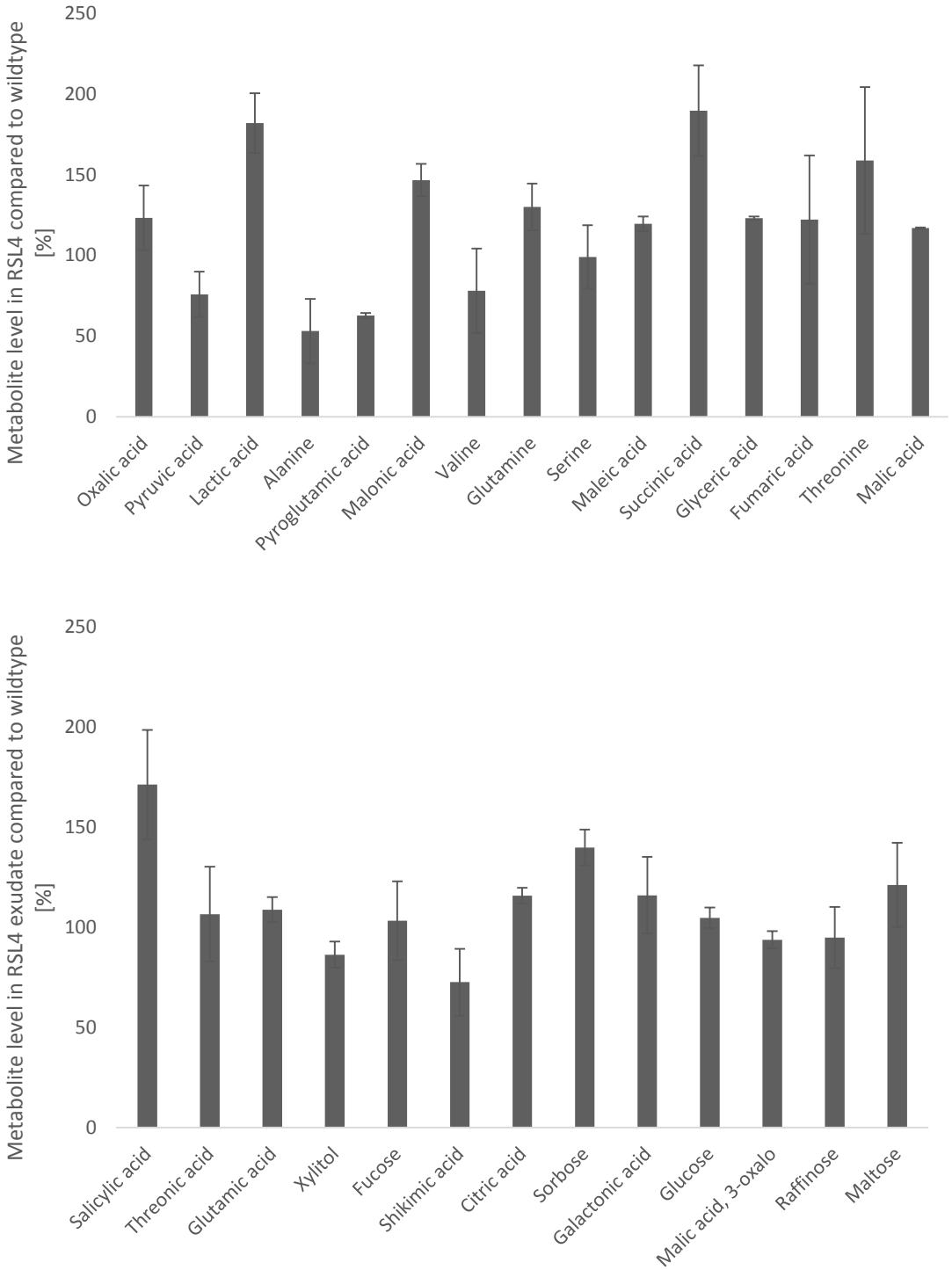


Figure 41: **Comparison metabolite levels in exudates from wildtype and RSL4-over-expressing seedlings.** Metabolite levels in exudates from RSL4 seedlings are expressed relative to those for Columbia

grown at the same time and under the same conditions. The absolute quantities of metabolites measured per gram fresh weight for each genotype are presented in Appendix 22. Data are displayed as the mean and standard deviation of fractional abundances measured from metabolites. A minimum of four samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets

7 Evaluating the feasibility of reliably measuring mass isotopomer distributions from specific cell populations in the roots of Arabidopsis thaliana seedlings

Chapter 3 demonstrated through *in silico* models that average data measured from a mixture of cells with different fluxes do not report on the actual fluxes in the system. Following from this Chapter 6 described investigation of stable isotope labelling of root hair cells to start the development of a model system for cell type specific metabolism. Most of the cell types within the root are not as easily isolated. Investigating the intracellular fluxes in these cells requires development of new strategies to complement the current standard techniques, which are not able to distinguish between metabolites from different cell types.

The previous chapters have focused on the validation of one such technique using GFP as a reporter protein for cell type specific metabolism. The low amount of protein purified from seedlings with cell type specific GFP expression remains the main challenge for the reporter protein strategy in Arabidopsis roots. In this chapter GFP immunopurified from four cell types (stele, endodermis, cortex and epidermis) is used to evaluate the feasibility of reliably measuring mass isotopomer distributions from specific cell populations in the roots of *Arabidopsis thaliana* seedlings in order to study cell type specific metabolism.

7.1 Introduction

The Arabidopsis root offers an excellent target for studies of root development and in recent years intense research has shed light on the molecular mechanisms behind differentiation and growth, cell patterning, branching of the root and responses to environmental signals. Work on the genetics that regulate the developing root is on-going and proving the root to be affected by a complex network of interactions (Petricka et al., 2012c).

A major part of the work on Arabidopsis roots has focused on identifying genes that play a role in the processes that are important for cell growth, differentiation and root development. As a result, several transcription factors and genes governing the patterning of the apical meristem have been identified. The apical meristem contains two pools of stem cells around the quiescent centre

comprised of mitotically less active cells at the tip of the root and protected by the root cap. Growth of the root is largely governed from here: the columella root cap arises from the apically positioned cells, whereas the basally and laterally positioned cells differentiate into the other cell types, including stele, endodermis, cortex and epidermis cells. Cells of different types arise from a single differentiated cell as linages that form one-cell thick tissue layers along the longitudinal axis of the root, with the cell type defined by the position of the cell (Dolan et al., 1993; Petricka et al., 2012c).

Studies of the differentiation process have uncovered several key transcription factors and genes, which affect the specification of cell types in the Arabidopsis root. As part of their characterisation, Arabidopsis lines with these cell type specific expression patterns of GFP have been created using the upstream noncoding sequence (promoter) either with or without the coding sequence attached (Lee et al., 2006). Four of these lines were selected for purification of GFP from specific cell types: expression of the SCARECROW transcription factor has a key role in maintaining the stem cell niche and in the ground tissue layer governs differentiation into endodermis and cortex (pSCARECROW::GFP; quiescent centre and endodermis (Birnbaum et al., 2003)), WOODEN LEG is a gene encoding a CRE cytokinin receptor that is required for differentiation into vasculature cells (pWOODENLEG::GFP; stele (Mahonen et al., 2000)), and WEREWOLF is part of a complex involved in determining non-root hair fate in the ground tissue (pWEREWOLF::GFP; epidermis (Lee and Schiefelbein, 1999; Sena et al., 2004)). In addition the pCORTEX::GFP line was used for expression in the cortex (Lee et al., 2006). The enhancer trap pET111 (Nawy et al., 2005) line with expression in the columella was also used for the initial tests for GFP purification and a line with RSL4 over expression (Chapter 6, (Yi et al., 2010)) and GFP expression in the root hairs was used to test the feasibility of purifying root hair specific GFP.

Despite the extensive work on the Arabidopsis root, similar obstacles to those facing metabolic flux analysis have slowed down research on different aspects of the different cell types. Studies based on isolating cells of a specific cell type for studies of transcriptomes and in a few cases variations in metabolite concentrations in the Arabidopsis root have been mostly based on fluorescence activated cell sorting (FACS). These investigations have resulted in a gene expression map of the root

((Birnbaum et al., 2003) and additionally for the quiescent centre (Nawy et al., 2005)), a map of protein expression (Petricka et al., 2012a), information on responses to different environmental and hormonal signals, such as salt stress and iron deprivation (Dinneny et al., 2008) and auxin responses (Bargmann et al., 2013), and measured differences in levels of some metabolites such as amino acids and dipeptides, glucosinolates and phenylpropanoids (Moussaieff et al., 2013; Petersson et al., 2015). In addition, metabolite levels have been probed with the aim to confirm significant differences in gene expression related to the secondary metabolism of some of the cell types have an effect on the metabolic activity of the cells (Rogers et al., 2012).

Together these investigations have only touched on metabolic processes, but suggest there are measurable differences in cell type specific metabolism and that they can be very difficult to capture and quantify without a deeper understanding of the intracellular metabolism. The challenges related to obtaining cell type specific information and small amounts of analytes present in the cells affect the work on specific cell types in all of these analysis types. However, the efforts towards a better understanding of cell type specific differences in all of these different layers of the complex network affecting plant metabolism offer an excellent framework for developing cell type specific metabolic flux analysis.

The major aim of this chapter is to probe the feasibility of reliably measuring mass isotopomer distributions from GFP with cell type specific expression (from stele, endodermis, cortex and epidermis) in order to investigate cell type specific metabolism of the cells within the Arabidopsis root. The specific objectives were to confirm GFP expression in the lines, purify GFP from the specific cell types and analyse the mass isotopomer distributions and fractional abundances of ^{13}C from reporter protein purified from Arabidopsis seedlings grown in 20% [$^{13}C_6$]-glucose. In addition to this, differences in gene expression between some of the lines were explored with a focus on variation that may have an effect on central metabolism. This chapter describes the advances made so far in using the reporter protein strategy to investigate cell type specific metabolism in the Arabidopsis root.

7.1.1 Gene expression related to central metabolism of specific cell types of the Arabidopsis root

As part of the recent work on populations of cells isolated with FACS, increasing amounts of gene expression data are becoming available for the different cell types in the Arabidopsis root. These data have been shown to contain differences between cell types, but the analysis has rarely focused on metabolism alone. A few examples of work that found differences in gene expression related to metabolism of the cell types under investigation have been described. Most notably, Rogers and colleagues found differences in levels of glycosylated flavonols that corresponded to a change in gene expression in the cortex cells (Rogers et al., 2012) and Lan and group describe differences in gene expression in genes closely related to central metabolism in root hair cells (Lan et al., 2013).

In this chapter the gene expression data for four different cell types is investigated with the goal of identifying any cell type specific significant variation that may directly affect central metabolism of the cells. A dataset with normalised data on stele, endodermis, columella and cortex cells amongst many other cell types was chosen based on its well reported use for building gene expression maps of the Arabidopsis root (Birnbaum et al., 2003; Brady et al., 2007). These data were collected using cell sorting that relies on the fluorescence of a GFP-marker expressed in the cells with the same promoter sequences as used for GFP expression in the work described here. The results agree with published work on the expression of genes related to central metabolism in the specific cell types: significant changes can be observed between cell types in processes such as amino acid metabolism and phenylpropanoid metabolism.

7.2 Results

7.2.1 Confirmation of cell type specific GFP expression

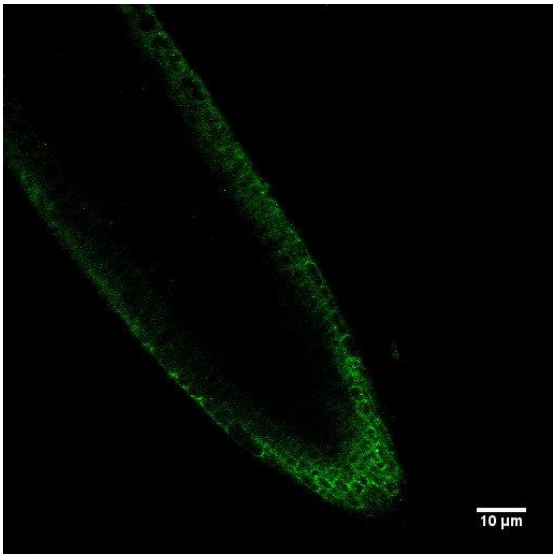
To confirm cell type specific GFP expression, seedlings of the four Arabidopsis lines with cell type specific GFP expression in stele, endodermis, cortex or epidermis were grown in hydroponic cultures for 14 days under similar conditions to those used in ^{13}C isotope labelling experiments. Confocal

images were acquired using a Zeiss LSM510 confocal microscope with propidium iodide used to stain the cells red to show the location of the GFP expression in the root.

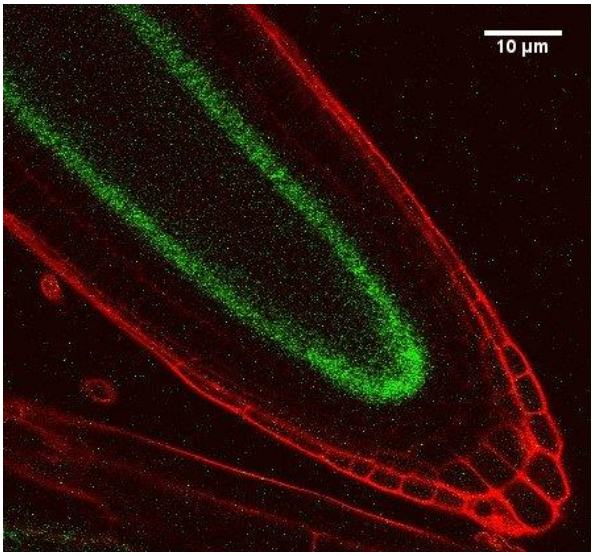
All four lines (pWEREWOLF::GFP, pSCARECROW::GFP, pCortex::GFP and pWOODENLEG::GFP) were found to have GFP expression (Figure 42) and the expression was observed in the cell types of interest. GFP expression was also detected in the pET111 line with expression in the columella, but it was clear the amounts of GFP in this line would be extremely low, due to the number of cells expressing GFP and the difficulties in imaging the expression. The expression of GFP in root hairs was observed, but was difficult to image from the roots. The line with expression in the root hairs contained clearly the lowest levels of GFP (data not shown).

While expression was detected from all lines, not all seedlings in the cultures showed GFP expression during imaging. This was especially true for the pET111 and pWOODENLEG lines. Ideally the amounts of GFP obtained from each culture should be maximised, but the only direct effect of no GFP expression in some of the seedlings is the decrease in amount of GFP purified from these cultures. Therefore lack of expression in some of the seedlings poses no direct obstacle for cell type specific metabolic flux experiments.

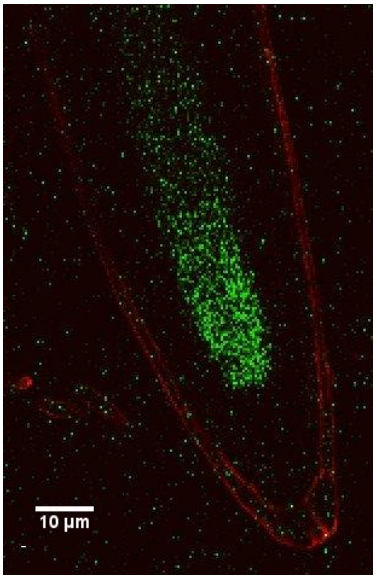
A pWEREWOLF::GFP



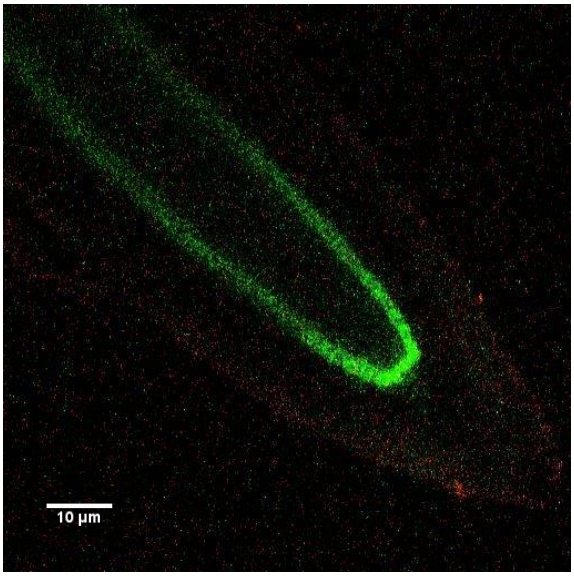
B pSCARECROW::GFP



C pWOODEN LEG::GFP



C pCORTEX::GFP



E pET111

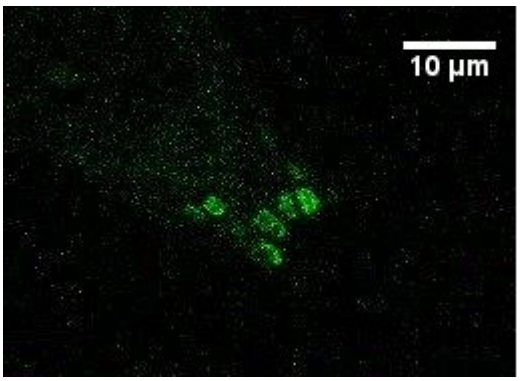


Figure 42: **Confocal images of transgenic plant lines with GFP expression in the expected cell types.** GFP expression (green) was visualised in roots of the following Arabidopsis lines: (A) pWEREWOLF::GFP, (B) pSCARECROW::GFP, (C) pWOODENLEG::GFP, (D) pCORTEX::GFP, and (E) pET111::GFP. In B and C cell walls were stained with propidium iodide to visualise the location of expression (red). Seedlings grown in hydroponic cultures for 14 days were removed from the cultures and some of the roots were excised, dipped in propidium iodide solution and washed with water before imaging. Scale bars 10 µm.

7.2.2 GC-MS detection of amino acids and mass isotopomer distributions derived from cell type specific GFP

GFP was purified from five Arabidopsis lines with cell type specific expression, in order to test the feasibility of detecting mass isotopomer distributions from the reporter protein. Fluorescence measurements were used to quantify the GFP purified from five cultures of 30-50 seedlings of each line (described in Chapter 4). The amount of GFP purified was between approximately 10 ng and 600 ng per culture of 30-50 seedlings. The highest amounts of GFP were purified from pCortex::GFP and pSCARECROW::GFP lines (200 - 600 ng per culture). The other lines showed similar variation but lower levels of GFP were purified: between 10 ng and 200 ng of GFP could be recovered from 30-50 seedlings from the pWEREWOLF::GFP and pWOODENLEG::GFP lines with less consistency between samples. The reasons for slightly higher variation in amounts compared to constitutive expression are likely to be two-fold: 1) confocal imaging revealed some of the roots did

not have GFP expression and 2) additional variation might have been added by loss of protein through purification steps. The latter might affect samples from lines with lower levels of expression more than samples where GFP is more abundant.

When analysed using standard methods and conventional protocols with the Agilent 5975C instrument, the total ion counts for amino acids detected from cell type specific GFP were below trace levels. However, these ion counts increased considerably, when the modified protocols used in Chapter 4 and 5 were used (Figure 43). The samples obtained from combining immunopurifed GFP from three cultures of 30-50 seedlings and using lower volumes of derivatisation reagents together with the two step purification produced levels high enough for detection of the amino acids, but still too low for reliable and reproducible detection of mass isotopomer distributions. The detected ion counts agreed well with the GFP expression observed in confocal images: the amino acid peaks detected from the pET111 line were considerably lower than those from the other lines.

To confirm the ion counts detected were too low for consistent and reliable analysis, fractional abundances of ^{13}C were calculated from the data for seedlings grown in 20% [$^{13}C_6$]-glucose. This analysis did show levels that were higher than the natural abundance of ^{13}C enrichment, but still did not report on the expected labelling provided in the uniformly labelled substrate. Variation was similar to that described for samples with too low ion counts in the analysis presented in Chapter 4, with a wide range of values both significantly above and below the expected ^{13}C enrichment of approximately 20% (Table 9). In addition, samples derived from seedlings grown in cultures with no additional isotope labelling added to the medium showed at least twice the natural abundance levels of ^{13}C enrichment for the majority of detected fragments, but also values close to zero. In some cases distinguishing between the two different types of samples would have been impossible based on the measured data and there was no clear pattern for data measured from different fragments of the same amino acid.

The modifications successfully used in Chapter 4 did therefore not achieve reliable detection of mass isotopomer data from amino acids derived from GFP from specific cell types. The overcome this

problem, the biological material used for the experiment was increased from three to five cultures of 30-50 seedlings. However, such low amounts of GFP were recovered from the pET111 line that work on it presented an additional challenge compared to the other lines with cell type specific GFP expression. It was excluded from the analysis until such point that it would be possible to obtain reliable mass isotopomer measurements from the other cell types with higher levels of expression. GFP derived from the other four Arabidopsis lines was immunopurified from five cultures and for each line, the cell type specific GFP was then combined into a single sample. This increase in biological material led to a significant difference in the total ion counts detected for amino acid fragments (Figure 44, Figure 45, Figure 46 and Figure 47 for each of the different cell types).

As with the constitutively expressed GFP, the ion counts for amino acids derived from the purified reporter protein varied between different samples of the same Arabidopsis line (Appendix 23) Since individual samples showed similar ion counts for amino acid fragments across the whole sample, the variation is likely to arise from variation in amounts of GFP purified from the plants and from loss of protein during the purification steps. The spectra with significantly lower levels of amino acid fragments were omitted from analysis of fractional abundances of ^{13}C. Where fractional abundances were calculated from these omitted data, they agreed with earlier results and did not show reliable detection of isotopic labelling.

For the fragments that were detected at higher ion counts (close to 100,000 or over), the calculated fractional abundances of ^{13}C for amino acid fragments gave approximately 20% labelling for GFP from seedlings grown in 20% [$^{13}C_6$]-glucose and the natural abundance of ^{13}C for GFP from seedlings grown in normal D-glucose. Exceptions to this were some fragments from the GFP from pCortex line. These fragments were detected at high levels, but contained a low fractional abundance of ^{13}C compared to the expected 20% provided in substrate or alternatively high fractional abundance (2-5%) in samples from seedlings that had grown in normal glucose. Many of these discrepancies involve the same amino acid fragments in both the labelled and natural abundance sample types, suggesting there could be an effect from peak overlap with contaminants. However, no source for cell type specific contamination could be identified in the experiment. All samples were grown,

purified and prepared the same way at the same time, and run in a random order. Therefore it is likely these fragments were still not detected at high enough levels to achieve consistent detection. Another option is that the variations in levels were not as significant as they appeared based on the ion counts, as they were not compared against internal standards.

The variation in ion counts meant that it was difficult to obtain sufficient measurements for a single amino acid fragment to allow a comprehensive comparison of the data. When the abundance of fragments was high enough, the detection was reliable and produced the expected labelling. Therefore increasing the number of samples run or increasing the number of cultures used for each sample remains an option that is likely to result in a dataset with enough data points for statistically significant comparisons. Both of these options increase the workload and cost of the experiments.

To avoid these obstacles, another option was to use a different analytical system with better sensitivity or larger injection volumes. In Chapter 4, the Agilent 7200 MS/Q-TOF instrument was able to measure significantly higher ion counts for amino acid fragments from standards. To test whether this increase in sensitivity would allow reliable mass isotopomer measurements from GFP purified from the lines with cell type specific expression, GFP was first purified from single cultures of 30-50 seedlings of each of the four lines. The samples were prepared using half of the conventional standard volumes of derivatisation reagents and the injection volume was set at 0.5 µl. Significantly higher total ion counts were detected with the Agilent 7200 system: the ion counts of amino acid fragments were similar to those obtained from three cultures with the 5975C mass spectrometer using 1 µl injections (Figure 48). Taking into account the differences in concentration and injection volume, the sensitivity was improved approximately 10-fold relative to that obtained with standard GC-MS. This is consistent with earlier results on standards (Chapter 4). The improved sensitivity suggests that if the experiments carried out with the standard GC-MS system were repeated on this system, the expected ion counts detected from the reporter protein would be similar to those routinely detected from total protein samples.

An experiment to test this hypothesis using GFP purified from three cultures of 30-50 seedlings with double the original injection volume was conducted using the Agilent 7200 instrument. A technical fault with instrument at the time of measurement resulted in the loss of the majority of amino acid peaks that elute at higher temperatures. Nevertheless, where the fractional abundances of ^{13}C could be calculated for amino acid fragments, approximately 20% fractional abundance was measured from fragments derived from cell type specific GFP purified from seedlings grown in 20% [$^{13}C_6$]-glucose. This was considered a good indication that it would be possible to measure mass isotopomer distributions. Using the Agilent 7200 mass spectrometer a fractional abundance of approximately 20% was also obtained for ^{13}C from amino acids derived from GFP expressed in root hairs and purified from five cultures of 30-50 seedlings (Figure 50).

However, when the experiment was repeated, the analysis was compromised by loss of protein during filter purification. The elution of protein from the filters for the small amounts of protein used frequently resulted in no elution, which was the main problem for analysis. The lack of filter purification on the other hand resulted in considerable differences in the spectra, with the contamination compromising detection of the amino acid peaks. While the data collected with the Agilent 7200 was of high quality with high total ion counts, the sample preparation ultimately limited the amount of data collected from the cell type specific samples within the time available. There is further capacity to increase injection volumes when using this system and large volume injection remains an option for improvement of detection for standard GC-MS systems by installation of a multimode inlet. The results demonstrate the feasibility of measuring ^{13}C fractional abundances from cell type specific GFP and show that future work may well lead to consistent and reliable detection of mass isotopomer distributions from this system.

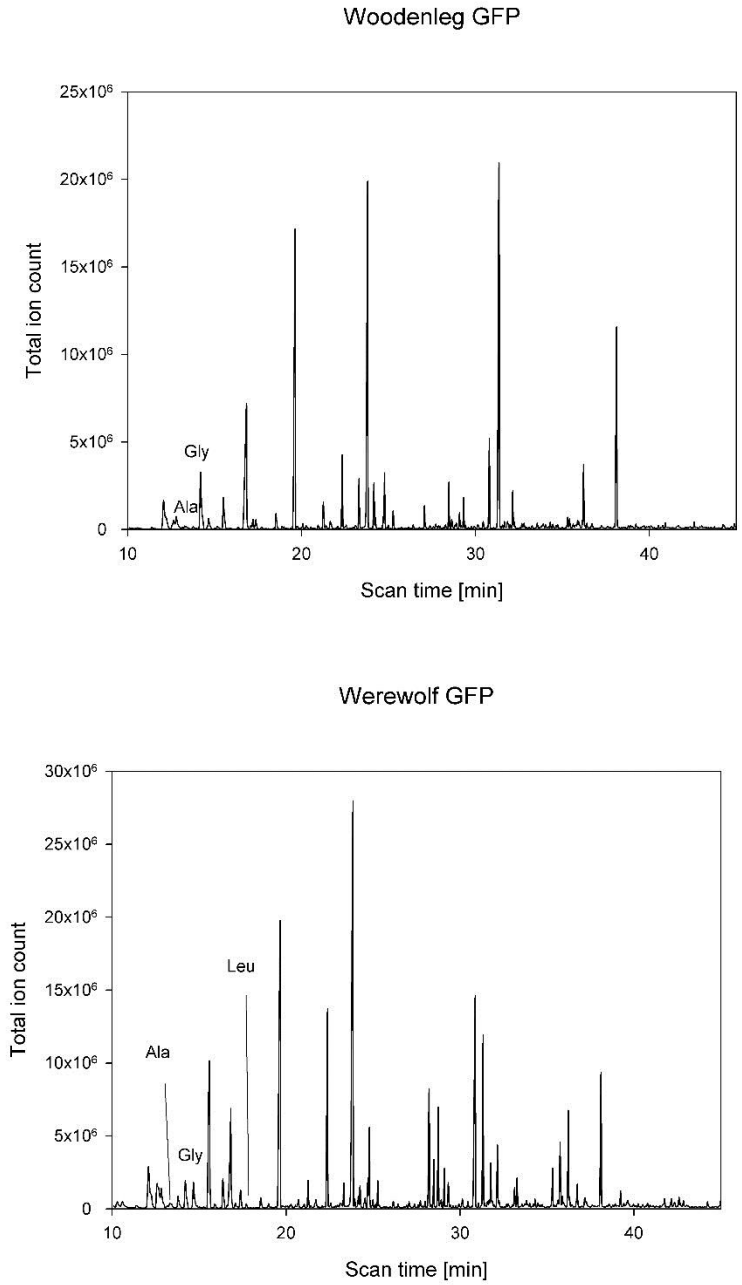
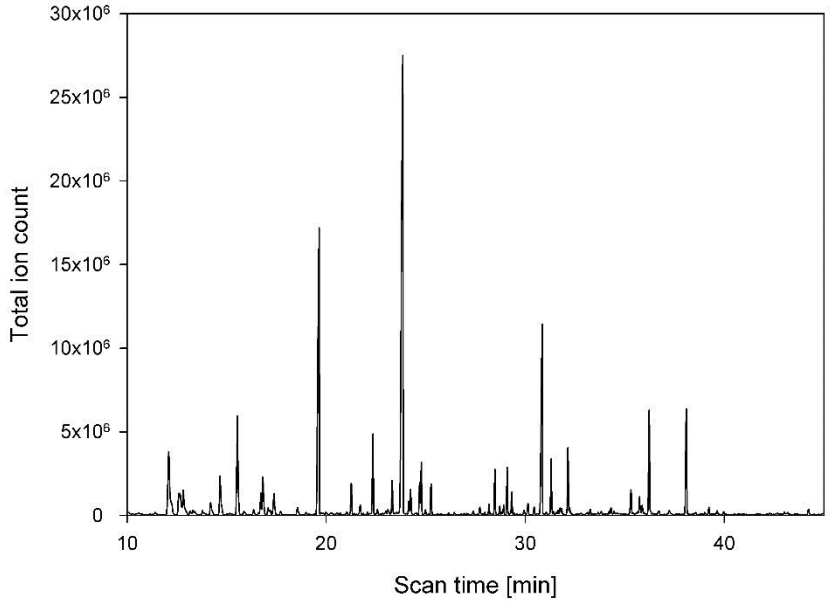


Figure 43: **Chromatographic detection of amino acids from GFP expressed in specific cell types.** GFP was immunopurified from three cultures of 30-50 seedlings for four lines with cell type specific expression. The purified protein was hydrolysed and derivatised with reduced total volumes of reagents (20 µl). Total ion chromatograms are shown for the four cell types. Peaks containing amino acid fragments are indicated, except for where the levels are too low to identify the corresponding peak from the chromatogram.

Scarecrow GFP



Cortex GFP

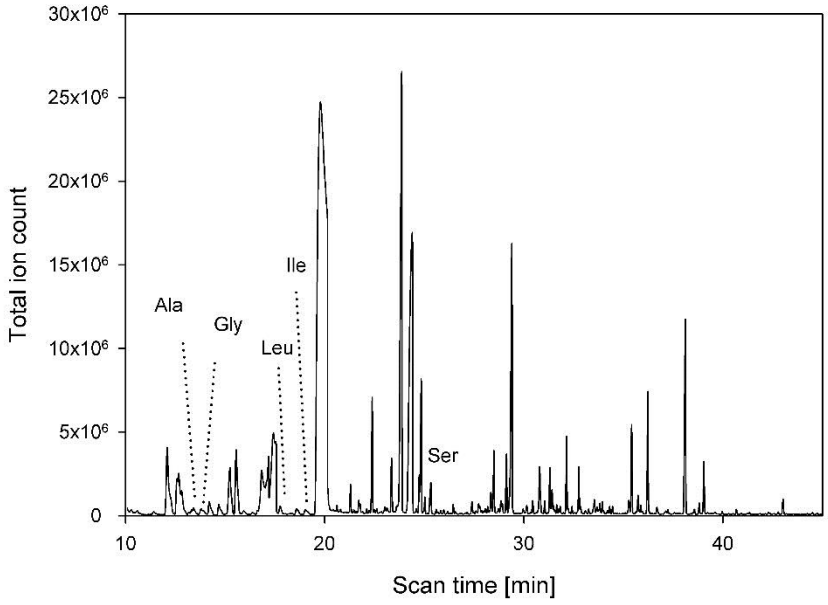
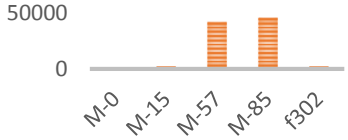
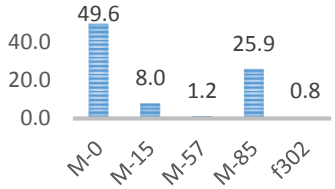
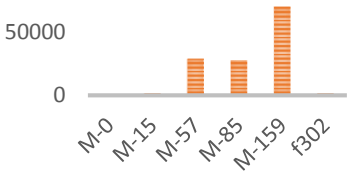
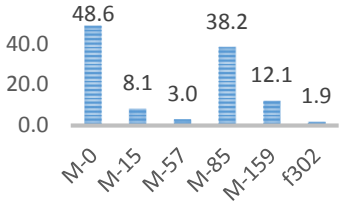
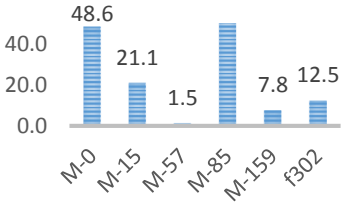
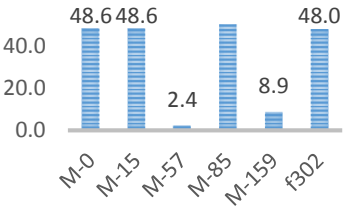
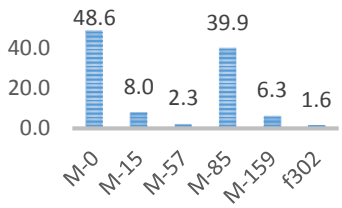


Table 9: **Fractional abundances of ^{13}C determined from alanine derived from GFP purified from three cultures of 30-50 seedlings from the lines with cell type specific expression and analysed using reduced volumes of derivatisation reagents.** Examples of fractional abundances calculated from amino acid fragment data derived from GFP purified from seedlings grown in either normal D-glucose or 20% [$^{13}C_6$]-glucose. Total ion counts for each fragment ion are presented and average fractional abundance calculated from the data are shown for each of these fragments.

Amino acid	Substrate	Total ion count	Average fractional abundance
Alanine	**Natural glucose**		
Alanine	**Natural glucose**		
Alanine	**20%-[$^{13}C_6$]glucose**		
Alanine	**20%-[$^{13}C_6$]glucose**		
Alanine	**20%-[$^{13}C_6$]glucose**		

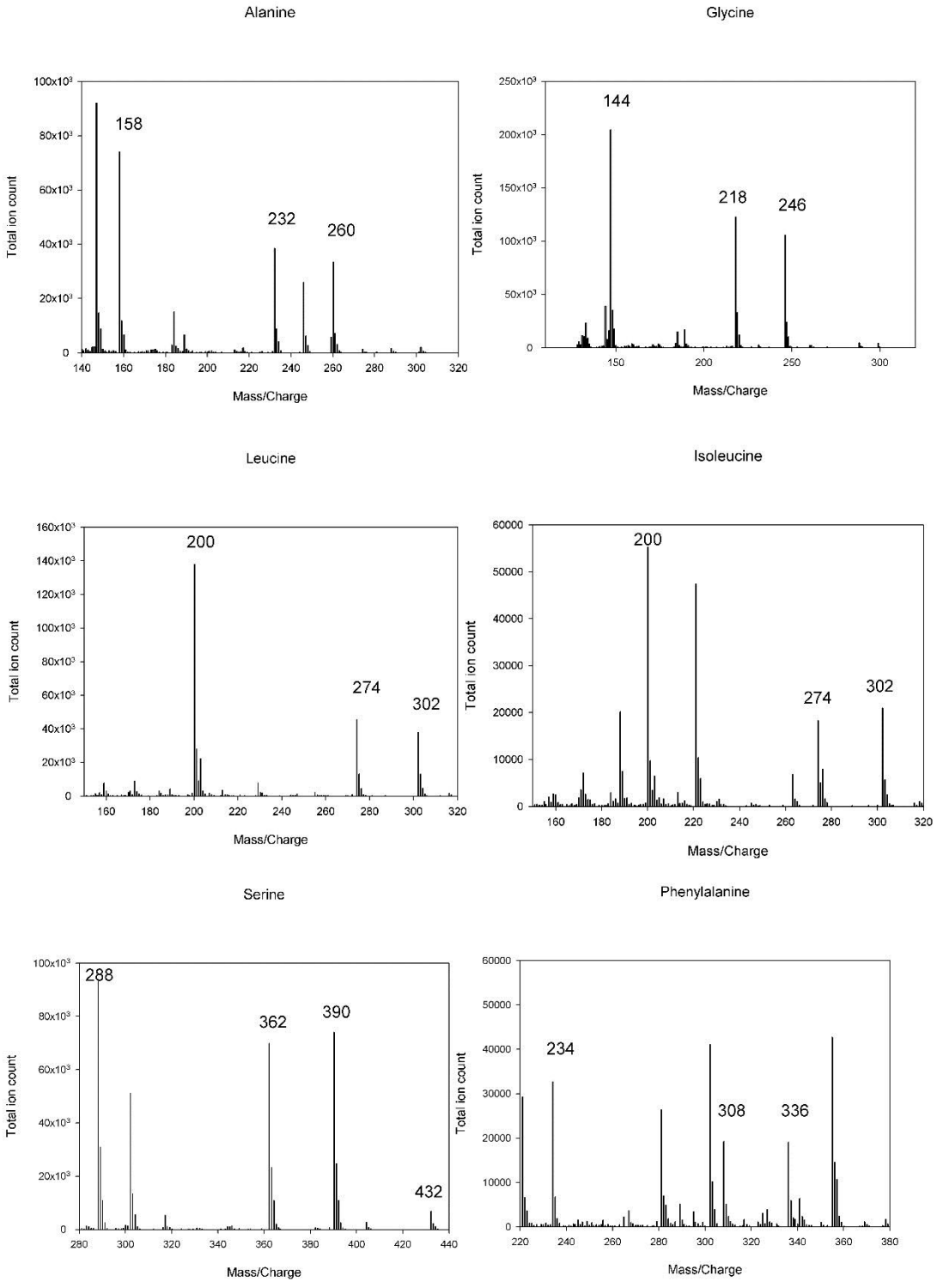
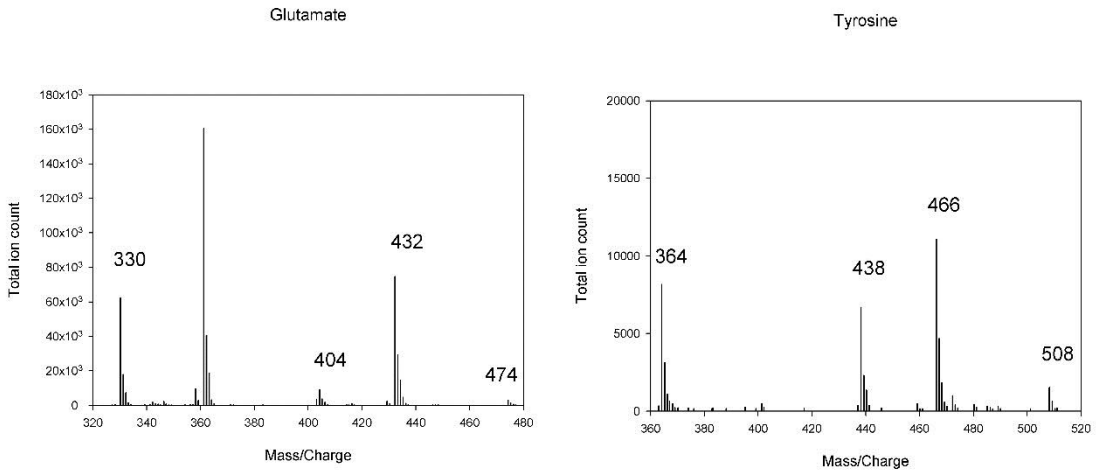


Figure 44: **Spectra of amino acid fragments derived from GFP purified from the pCortex::GFP line from five cultures of 30-50 Arabidopsis seedlings using an Agilent 5975C mass spectrometer and reduced volumes of derivatisation reagents.** Scale for each spectrum has been adjusted to show all fragments detected.


Glutamate
Tyrosine
Total ion count
Mass/Charge
330
404
432
474
364
438
466
508

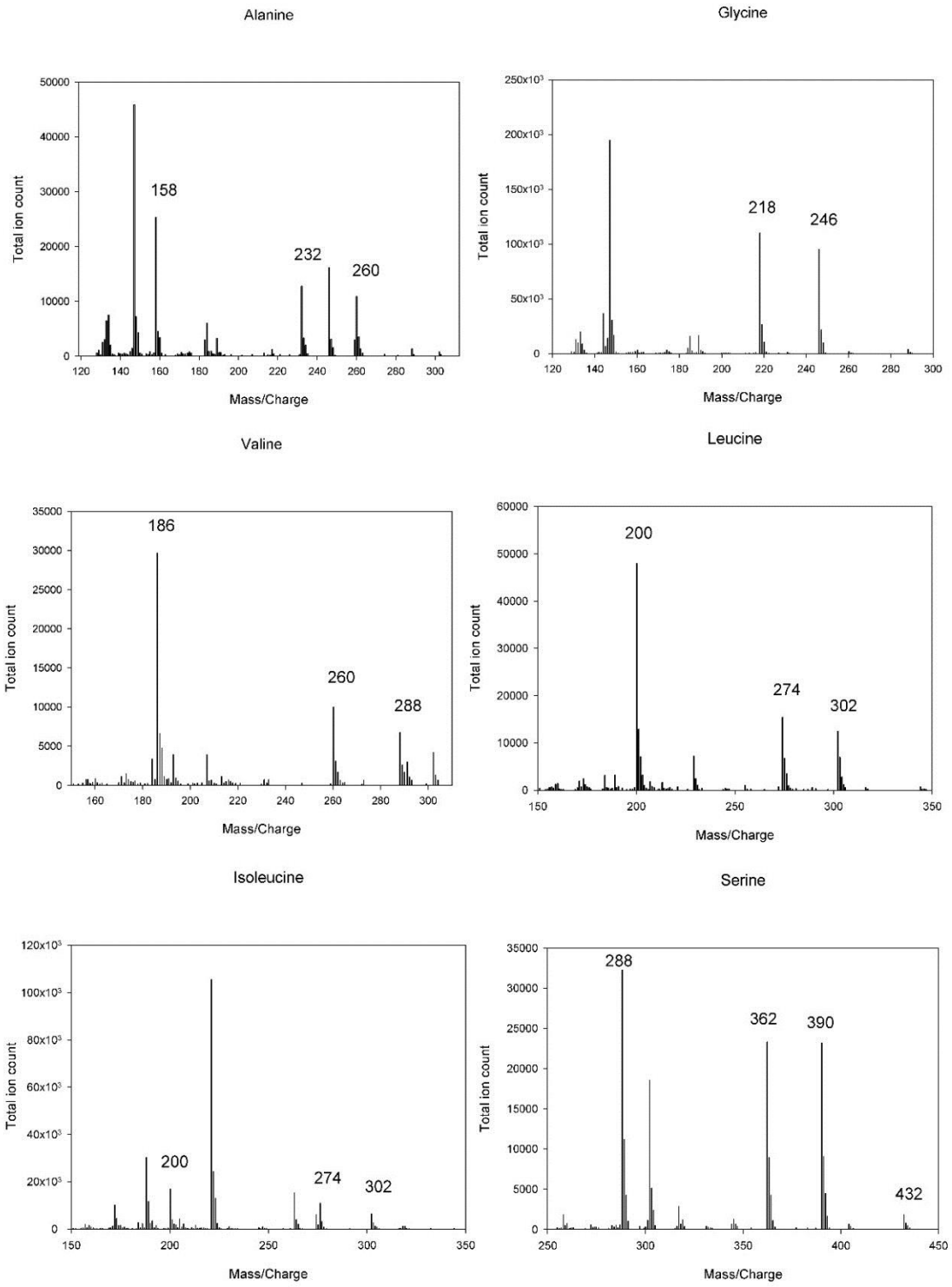
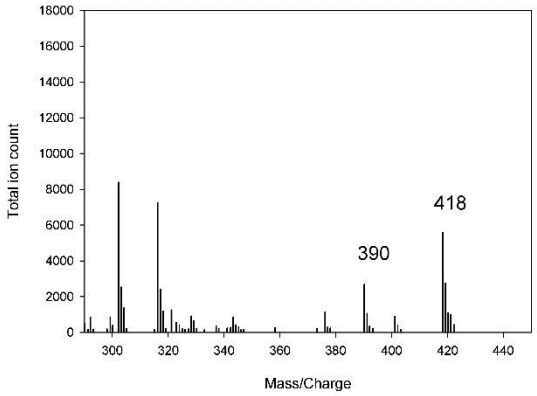
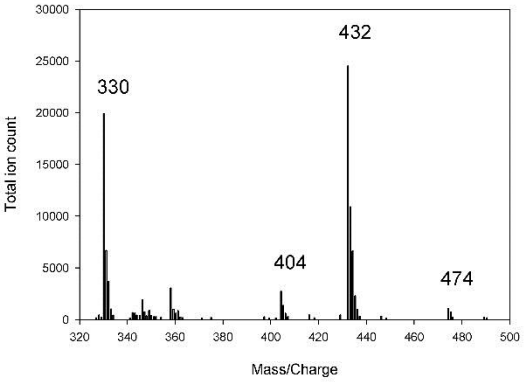


Figure 45: **Spectra of amino acid fragments derived from GFP purified from the pSCARECROW::GFP line from five cultures of 30-50 Arabidopsis seedlings using an Agilent 5975C mass spectrometer and reduced volumes of derivatisation reagents.** Scale for each spectrum has been adjusted to show all fragments detected.

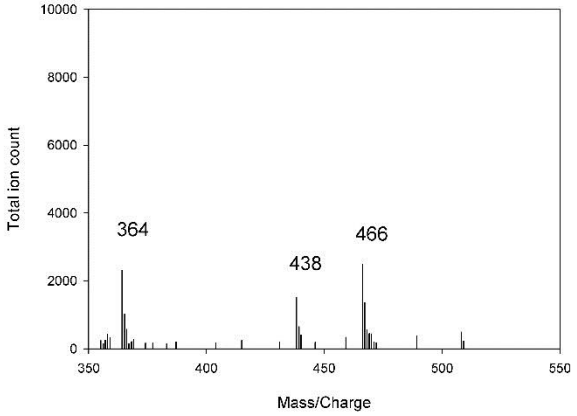
Aspartate



Glutamate



Tyrosine



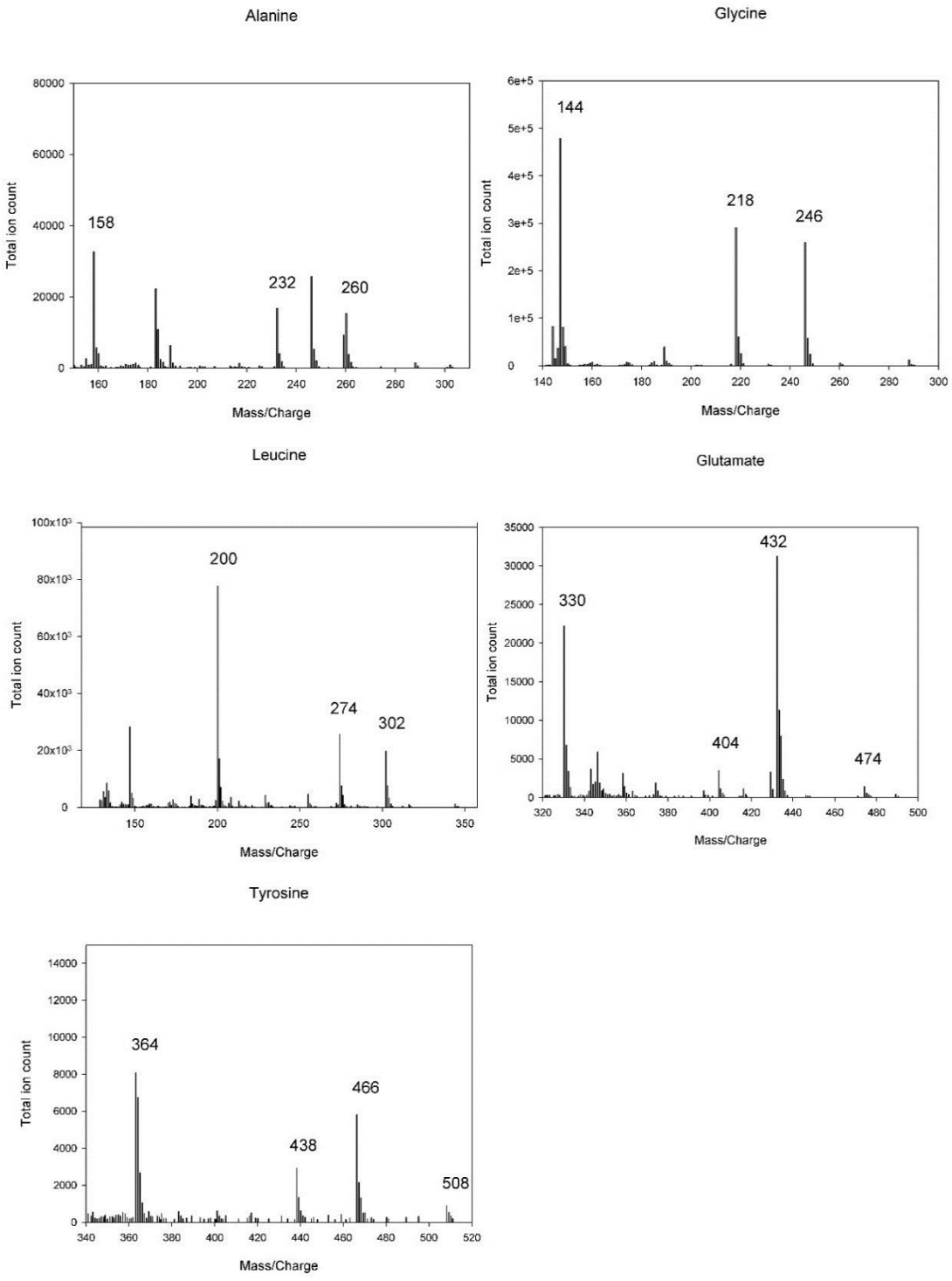


Figure 46: **Spectra of amino acid fragments derived from GFP purified from the pWEREWOLF::GFP line from five cultures of 30-50 Arabidopsis seedlings using an Agilent 5975C mass spectrometer and reduced volumes of derivatisation reagents.** Scale for each spectrum has been adjusted to show all fragments detected.

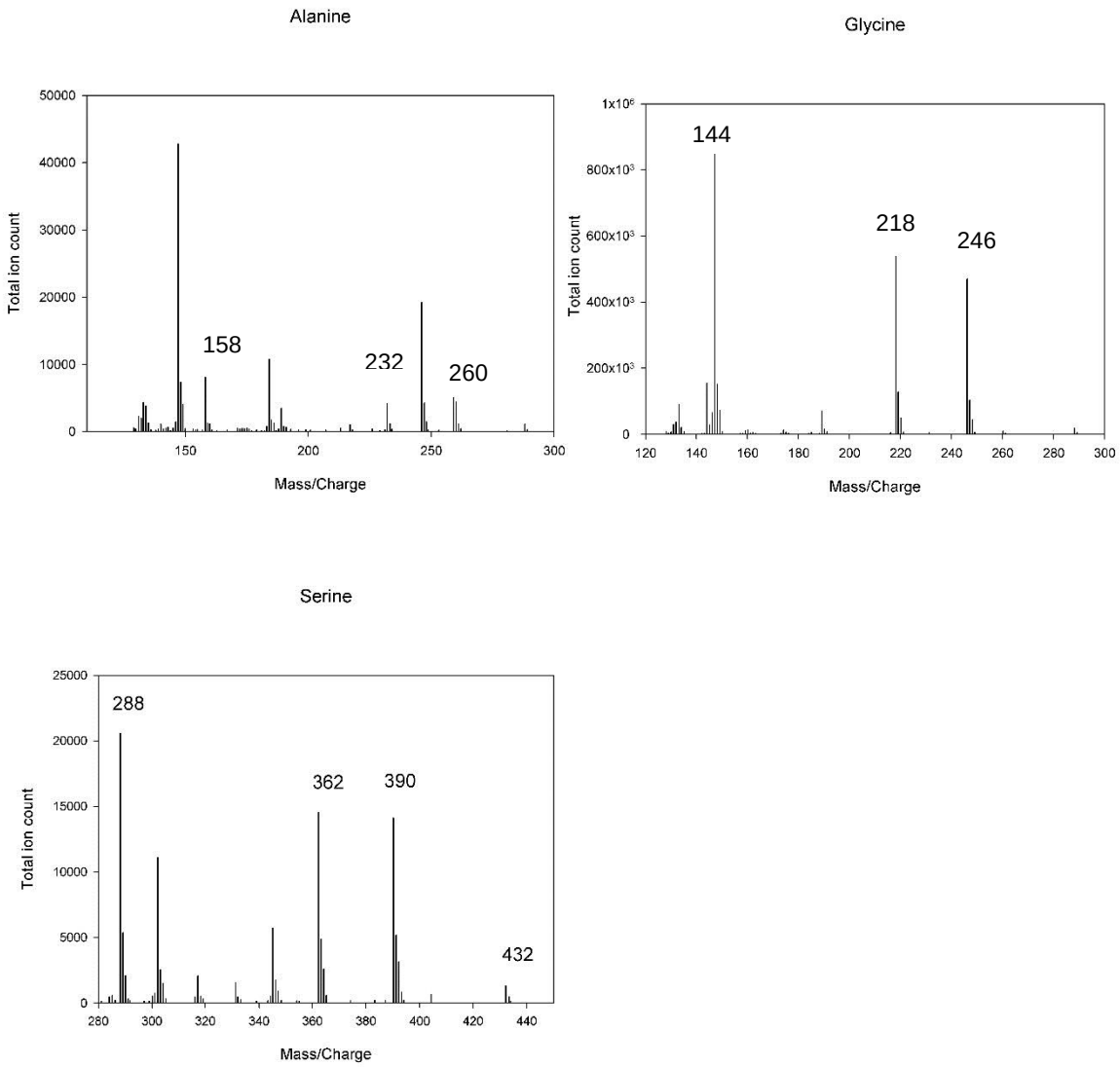
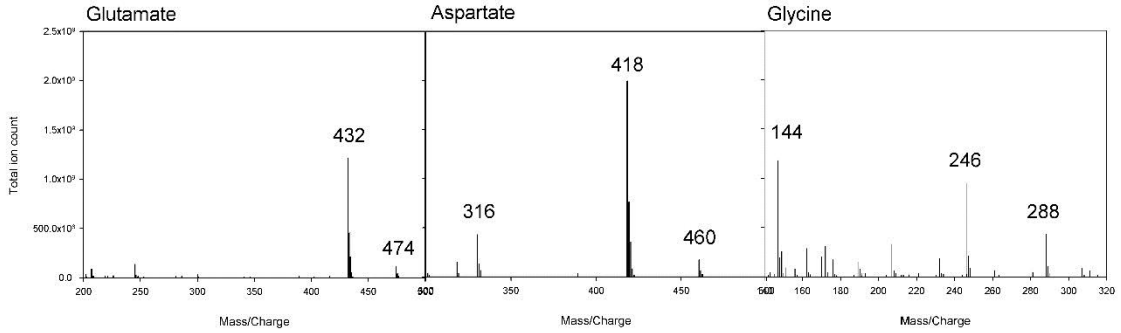
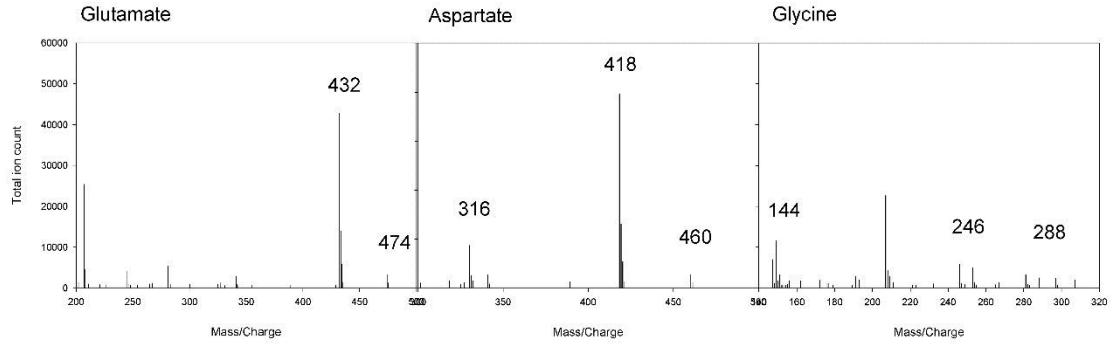


Figure 47: **Spectra of amino acid fragments derived from GFP purified from the pWOODENLEG::GFP line from five cultures of 30-50 Arabidopsis seedlings using an Agilent 5975C mass spectrometer and reduced volumes of derivatisation reagents.** Scale for each spectrum has been adjusted to show all fragments detected.

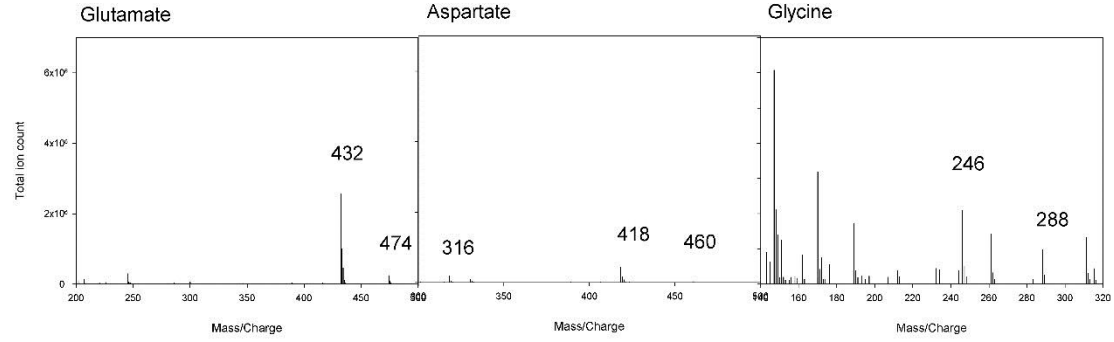
A) pCortex::GFP



D) pWEREWOLF::GFP



C) Constitutive expression



B) pSCARECROW::GFP

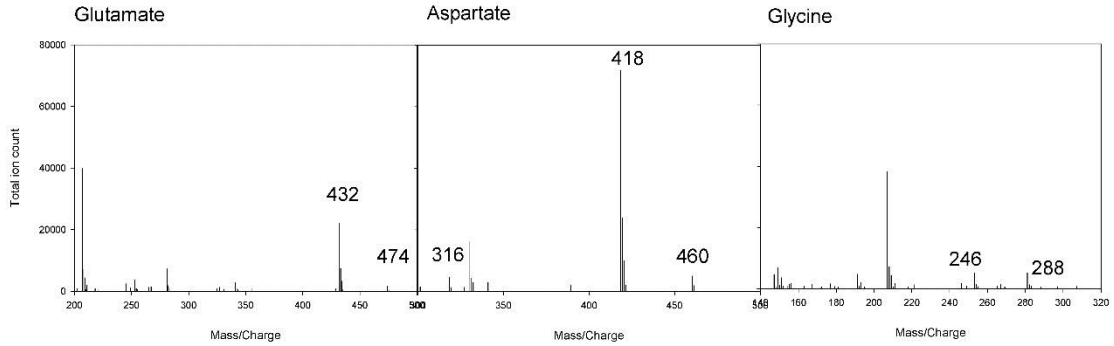


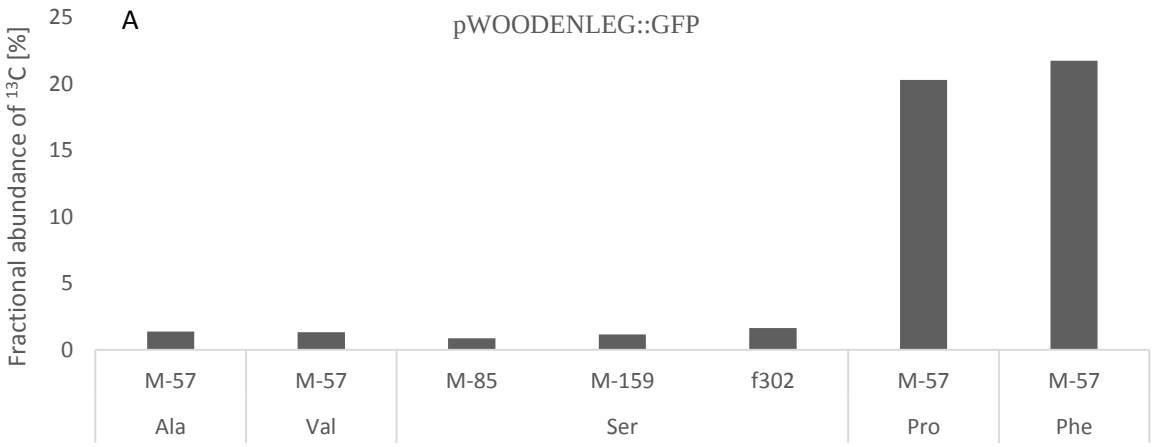
Figure 48: **Spectra of amino acid fragments from cell type specific GFP purified from single cultures of 30-50 Arabidopsis seedlings analysed on an Agilent 7200 mass spectrometer.** Injection volume for the experiment was half of normal (0.5 µl) and half of the normal total volume of derivatisation reagents was used. Mass spectra for glycine, aspartate and glutamate fragments are

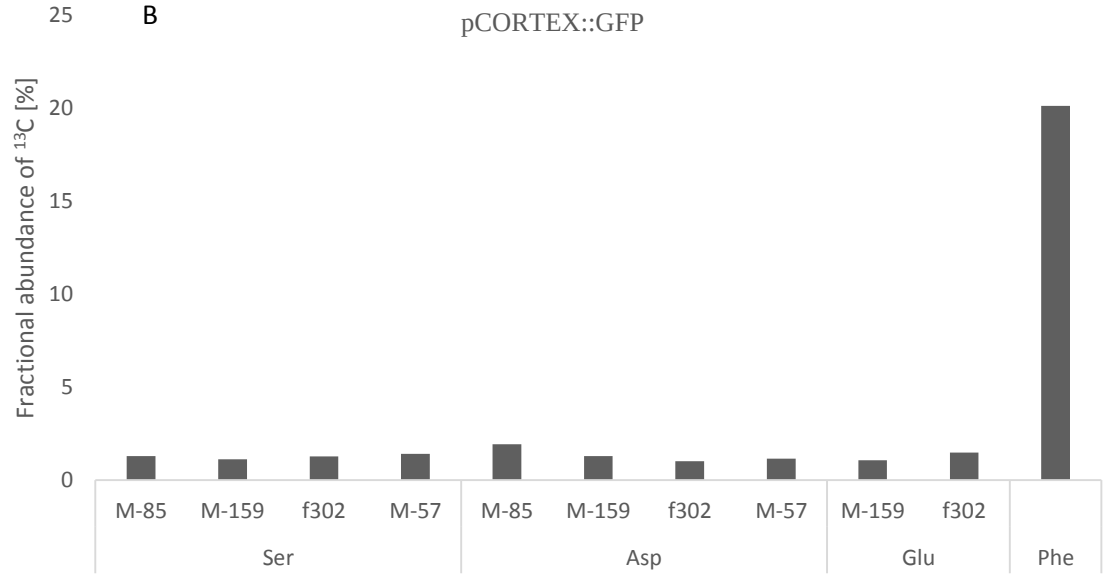
shown for A) pCortex::GFP, B) pWEREWOLF::GFP, C) GFP purified from the line with constitutive expression, and D) pSCARECROW::GFP.

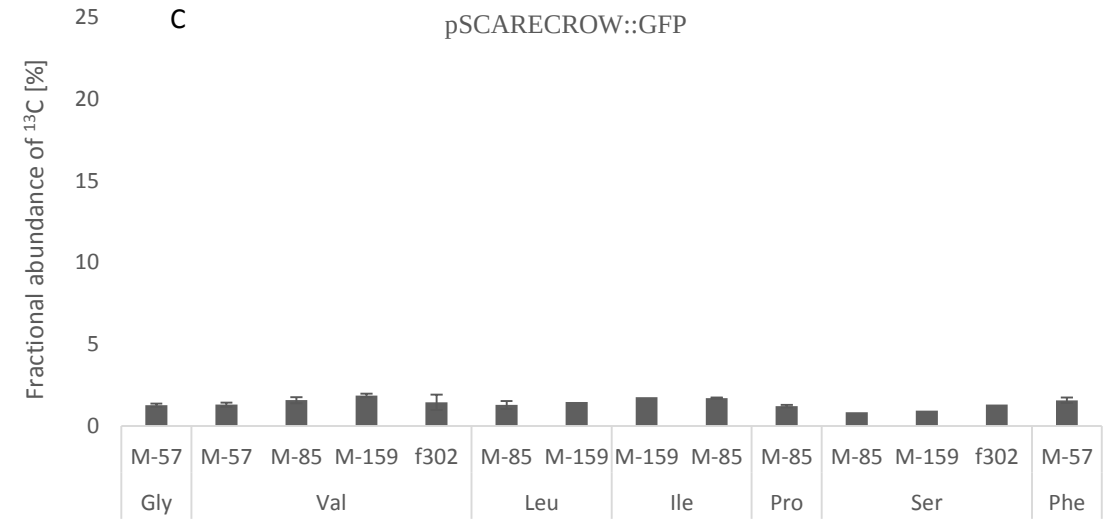
7.2.3 Average fractional abundance of ^{13}C detected from specific cell types in the Arabidopsis root

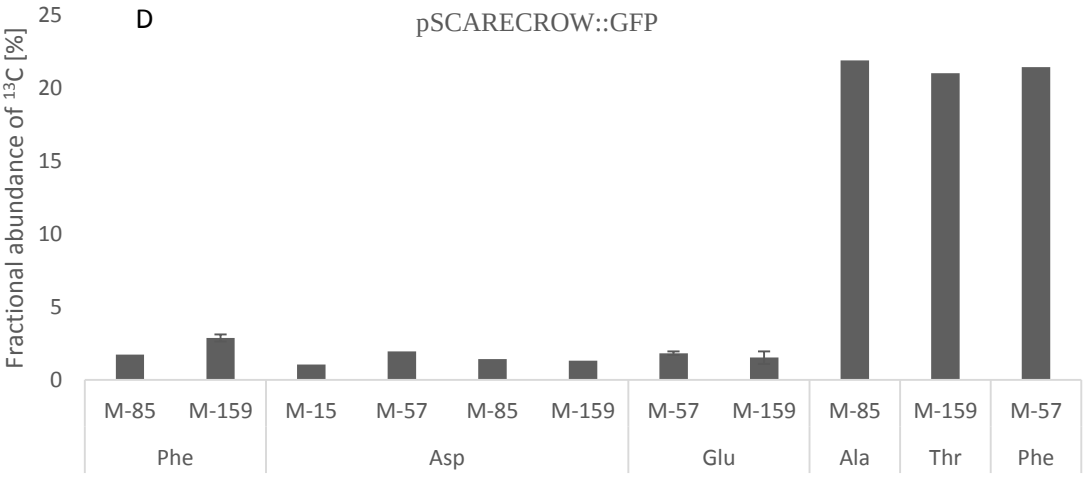
The detection of mass isotopomer distributions in amino acid fragments was not achieved consistently with the standard GC-MS system. Nevertheless, the experiment with GFP purified from five cultures of 30-50 seedlings resulted in sets of data that show natural abundance of ^{13}C in several fragments from GFP purified from seedlings of the different Arabidopsis lines with cell type specific GFP expression grown in D-glucose. When grown in 20% [$^{13}C_6$]-glucose, the fractional abundances of ^{13}C of several fragments reached approximately 20% (Figure 49). An exception to this was the pCortex line, where fractional abundances were approximately 9-15% for several of the amino acid fragments detected at highest ion counts.

Despite limited measurements with the 7200 mass spectrometer, the fractional abundances of ^{13}C for some fragments such as glycine were calculated from the data (Figure 50). These showed approximately 20% ^{13}C enrichment for all of the lines, including a sample of GFP from the root hairs. The data collected with the 7200 mass spectrometer does not agree with the measurements from pCortex line that showed a reduction in fractional abundances in fragments from alanine when the 5975C mass spectrometer was used (data not shown). It is therefore very likely the measurements made even at seemingly high levels using GFP purified from five cultures and measured using the 5975C mass spectrometer still were below the detection limit for reliably detecting mass isotopomer distributions from amino acid fragments. The increase in total ion counts achieved with the


A
pWOODENLEG::GFP
Fractional abundance of 13C [%]
25
20
15
10
5
0
M-57
Ala
M-57
Val
M-85
M-159
Ser
f302
M-57
Pro
M-57
Phe


B
pCORTEX::GFP
Fractional abundance of 13C [%]
25
20
15
10
5
0
M-85
M-159
f302
M-57
Ser
M-85
M-159
f302
M-57
Asp
M-159
f302
Glu
Phe


C
pSCARECROW::GFP
Fractional abundance of 13C [%]
25
20
15
10
5
0
M-57
M-57
M-85
M-159
f302
M-85
M-159
M-159
M-85
M-85
M-85
M-159
f302
M-57
Gly
Val
Leu
Ile
Pro
Ser
Phe


D
pSCARECROW::GFP
Fractional abundance of 13C [%]
25
20
15
10
5
0
M-85
M-159
M-15
M-57
M-85
M-159
M-57
M-159
M-85
M-159
M-57
Phe
Asp
Glu
Ala
Thr
Phe

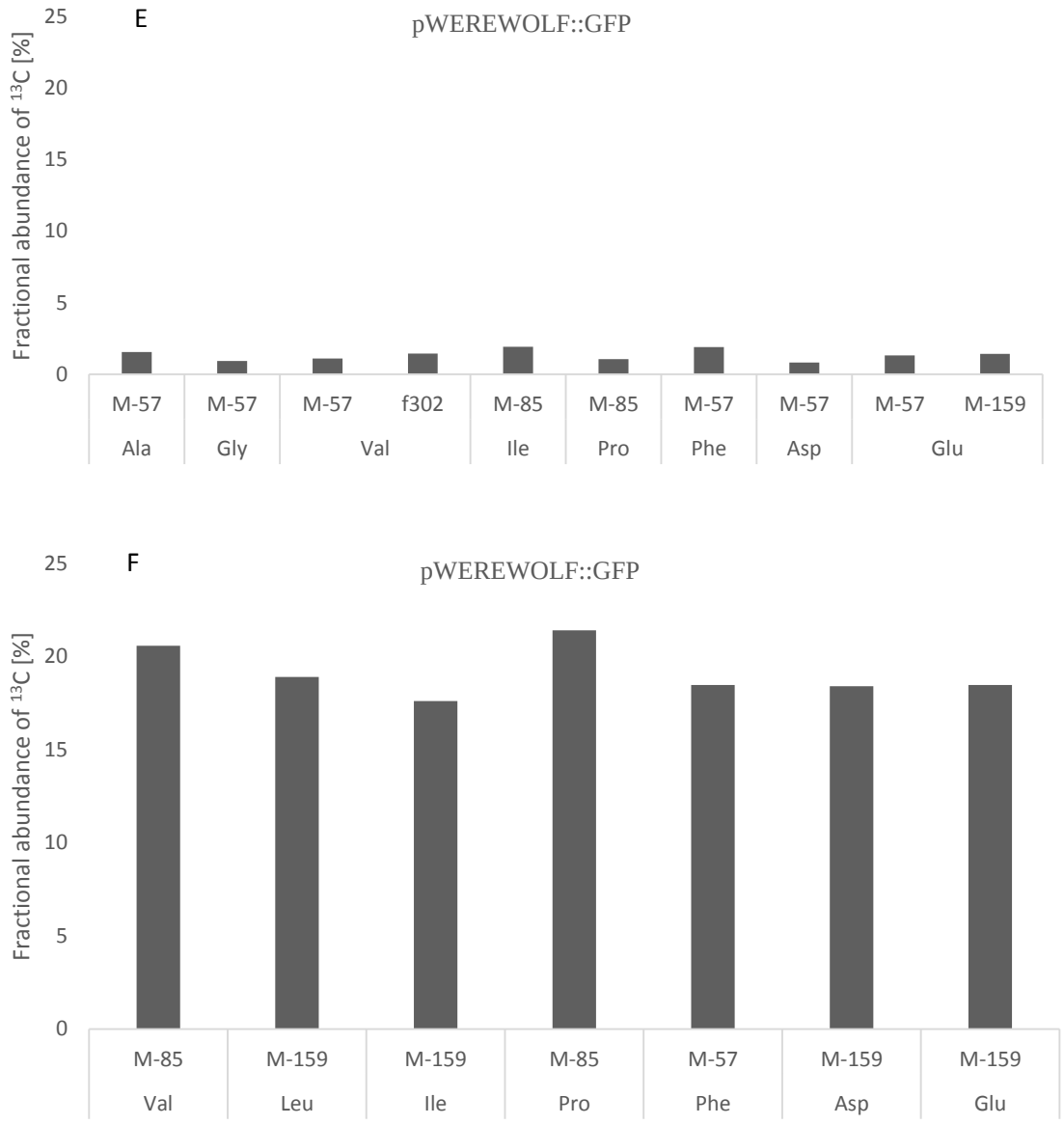


Figure 49: **Fractional abundance of ^{13}C in amino acid fragments detected from GFP purified from four different Arabidopsis lines with cell type specific GFP expression and detected with Agilent 5975C GC-MS.** Seedlings were grown either in normal glucose or 20% [$^{13}C_6$]-glucose. GFP purified from five cultures of 30-50 seedlings was combined and reduced volumes of derivatisation reagents were used for the analysis (20 µl). With the exception of fragments from the pCortex line (data not shown), all fragments are showing approximately the expected level of ^{13}C enrichment (either natural abundance or 20%), but only fragments at high ion counts were chosen for analysis. Standard deviations are shown where the number of samples was three or more. In figures A), B) and D) fragments from both groups of labelling are shown. A) pWOODENLEG::GFP, seedlings grown in natural glucose and 20% [$^{13}C_6$]-glucose, B) pCORTEX::GFP, seedlings grown in natural glucose and 20% [$^{13}C_6$]-glucose, C) pSCARECROW::GFP, seedlings grown in natural glucose, D) pSCARECROW::GFP, seedlings grown in natural glucose and 20% [$^{13}C_6$]-glucose, E) pWEREWOLF::GFP, seedlings grown in natural glucose, and F) pWEREWOLF::GFP, seedlings grown in 20% [$^{13}C_6$]-glucose.

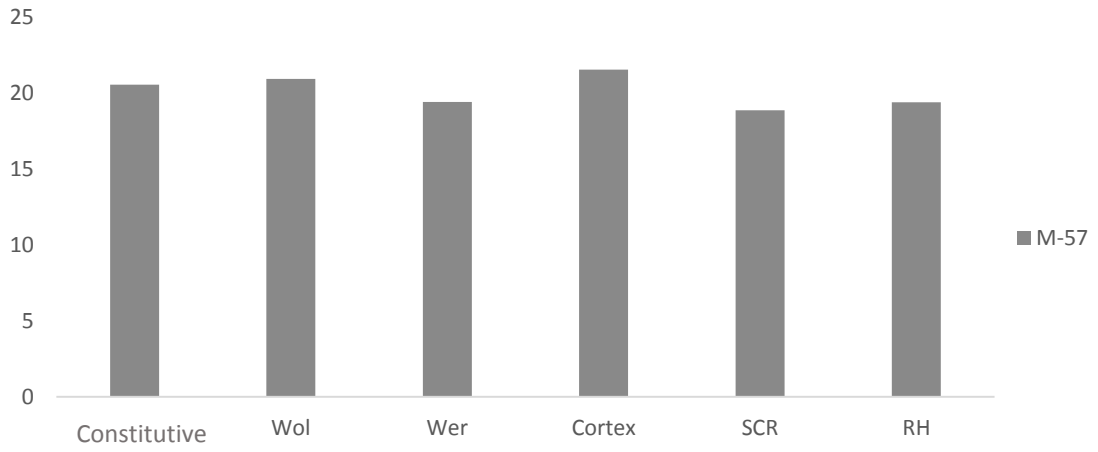


Figure 50: **Fractional abundances of ^{13}C detected from cell type specific GFP purified from three cultures of 30-50 seedlings grown in 20% [$^{13}C_6$]-glucose and detected with Agilent 7200 from glycine M-57 fragment.** Injection volume was increased to 1 µl. RH represents GFP purified from root hairs, Wol pWOODENLEG::GFP, Wer pWEREWOLF::GFP, Cortex pCORTEX::GFP, and SCR pSCARECROW::GFP.

7.2.4 Analysis of gene expression to identify changes directly related to central carbon metabolism

Gene expression patterns between whole root data and different cell types including cortex, endodermis, stele, columella and epidermis (Thimm et al., 2004) were compared using MapMan software. The software allows visual mapping of the data on metabolic pathways based on the genome provided by the Arabidopsis Information Resource (TAIR9 Genome Release). MapMan was set to use Wilcoxon rank sum test and confirm p-values using Benjamini-Hochberg procedure to control for false discovery rate. The resulting significant values under the threshold (here the conventional 0.05 significance threshold was used) were compared against the proportion of significant p-values in the whole dataset of gene expression data, producing the probability for seeing

the number of significant p-values observed in the sub set of genes in the whole set of data. The gene expression patterns are shown in the analysis as a false colour heat map (Usadel et al., 2009).

The MapMan software uses a scavenger module that classifies the genes in an array into "bins" based on their hierarchical functional category (Usadel et al., 2005). This contains information on the main bin, sub-bins and enzymes and is used to map the gene expression data onto the visual pathways presented through the user interface. Wilcoxon rank sum test in MapMan compares the variation in the expression of genes in a specific bin to all the other bins. This process can reveal statistically significant correlations between several genes connected to a single function. The variation in expression levels of expression of single genes may in this case be small, but the variation in the genes in the bin may still be highly significant. An example of this was presented by Usadel and colleagues for metabolic reactions in the bin "photosystem.light reaction" as night and day end: only small differences are observed in individual genes and they do not appear highly significant, while the response of the total bin gives the extremely small probability of 5.68×10^{-12} for being identical to the overall distribution (Usadel et al., 2005).

The gene expression in the different cell types from Arabidopsis roots was compared both for significant variation in the expression levels of individual genes related to central metabolism and for the significant response related to the functional hierarchical categories represented by bins. Data from each of the gene expression arrays for cells from the stele, endodermis, cortex, columella and epidermis were first compared against the whole root data. Parts of the individual heat maps depicting the gene expression profiles of each of the cell populations were then compared for differences between the cell types.

Using TCA cycle as an example, some genes do show significant differences in their expression patterns between the cell types (Figure 51). Most notable are differences in expression of some of the genes encoding isocitrate dehydrogenases. More broadly, the analysis suggests there might generally may be a higher level of expression of genes related to the TCA cycle in the cell type specific data. To investigate whether this over expression pattern was significant, Wilcoxon rank

sum test results for the bins containing these genes were collated for each cell type. For the steps that showed most variation between cell types visually, the variation did was not significant based on the p-values. However, two cases of over expression were found to be significant within the TCA cycle: one was the over expression of genes encoding aconitate hydratases (citrate to isocitrate) which was significant in all cell types, the other was genes encoding succinyl-CoA ligases (succinyl-CoA to succinate) in the epidermis which were significantly over expressed relative to the whole root only in the epidermis. Comparison between cell types reveals individual genes additionally vary in their expression in some instances where the variation in the gene functional group (bin) to which they belong was not found significant (Figure 52).

To probe the gene expression related to the metabolism of the cell types further, the data for each cell type was mapped on metabolic pathways that contained other key reactions that would be central for metabolic flux analysis. These included the pathways directly connected to the TCA cycle, such as gluconeogenesis and mitochondrial electron transport. Figure 53 shows a map of gene expression for each of the cell types. The most distinctive differences that arise in these patterns are again related to steps, where a large number of genes encode enzymes that take part in the same enzymatic step. These include genes encoding for pyruvate dehydrogenases (pyruvate to acetyl-CoA) and isocitrate dehydrogenases (isocitrate to alpha-oxoglutarate). For the pyruvate dehydrogenase genes, all of the cell types show significant over expression for at least one group of genes related to this enzymatic step, with significant variation in a larger number of genes for endodermis. The bins covering 93 genes related to mitochondrial electron transport show significant variation ($p<0.0003$) in all cell types. The step between citrate and oxaloacetate (ATP citrate lyase) is also highlighted as significantly different from the whole distribution in all cell types in addition to the variation related to aconitases described above. Cell type specific patterns include the significantly varied succinyl-CoA ligase in epidermis cells and variation in genes related to malic enzyme in columella. For genes in other bins, the p-values do not suggest the variation differs from normal variation within the data.

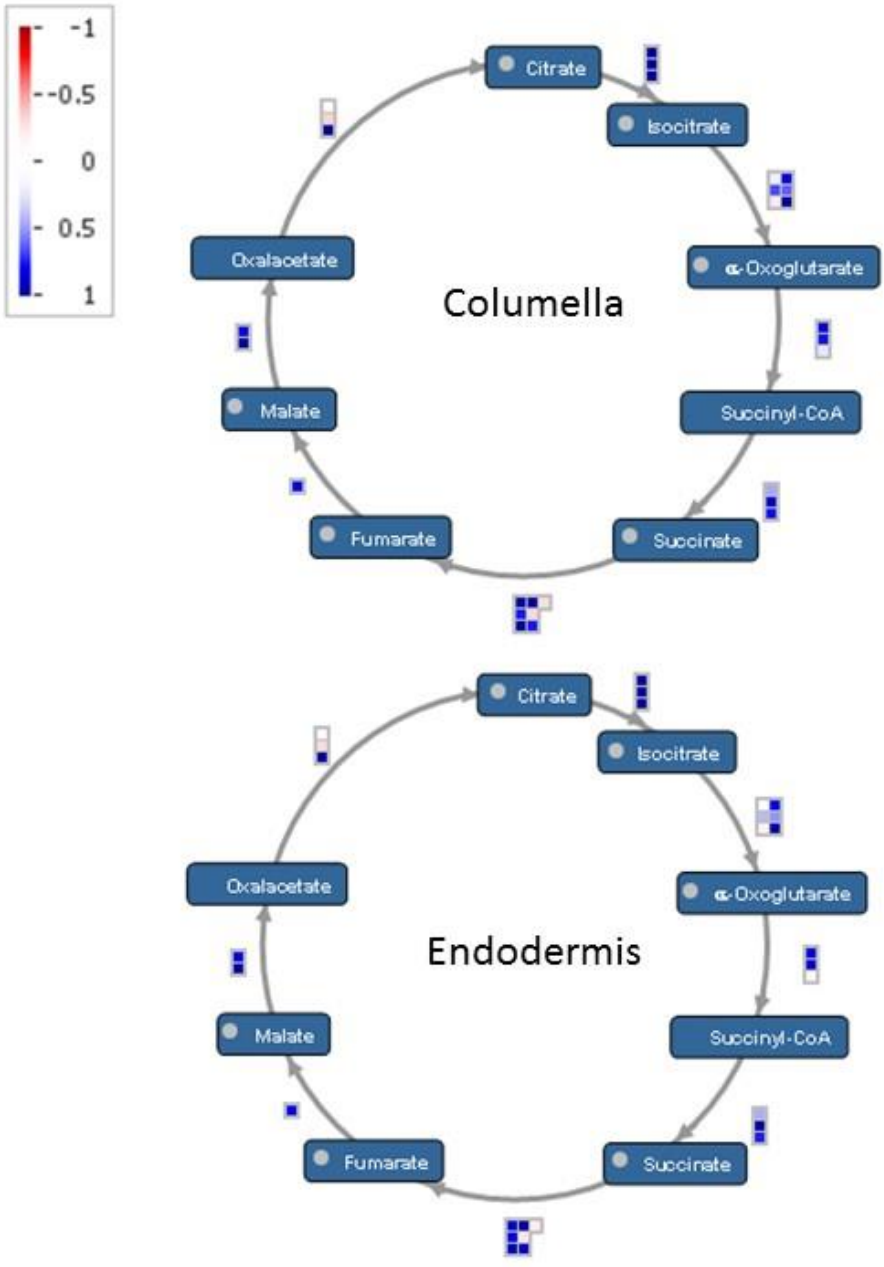
This initial analysis suggests gene expression data show similar profiles for the genes involved in the key processes in central metabolism in the different cell types. However, more varied significant

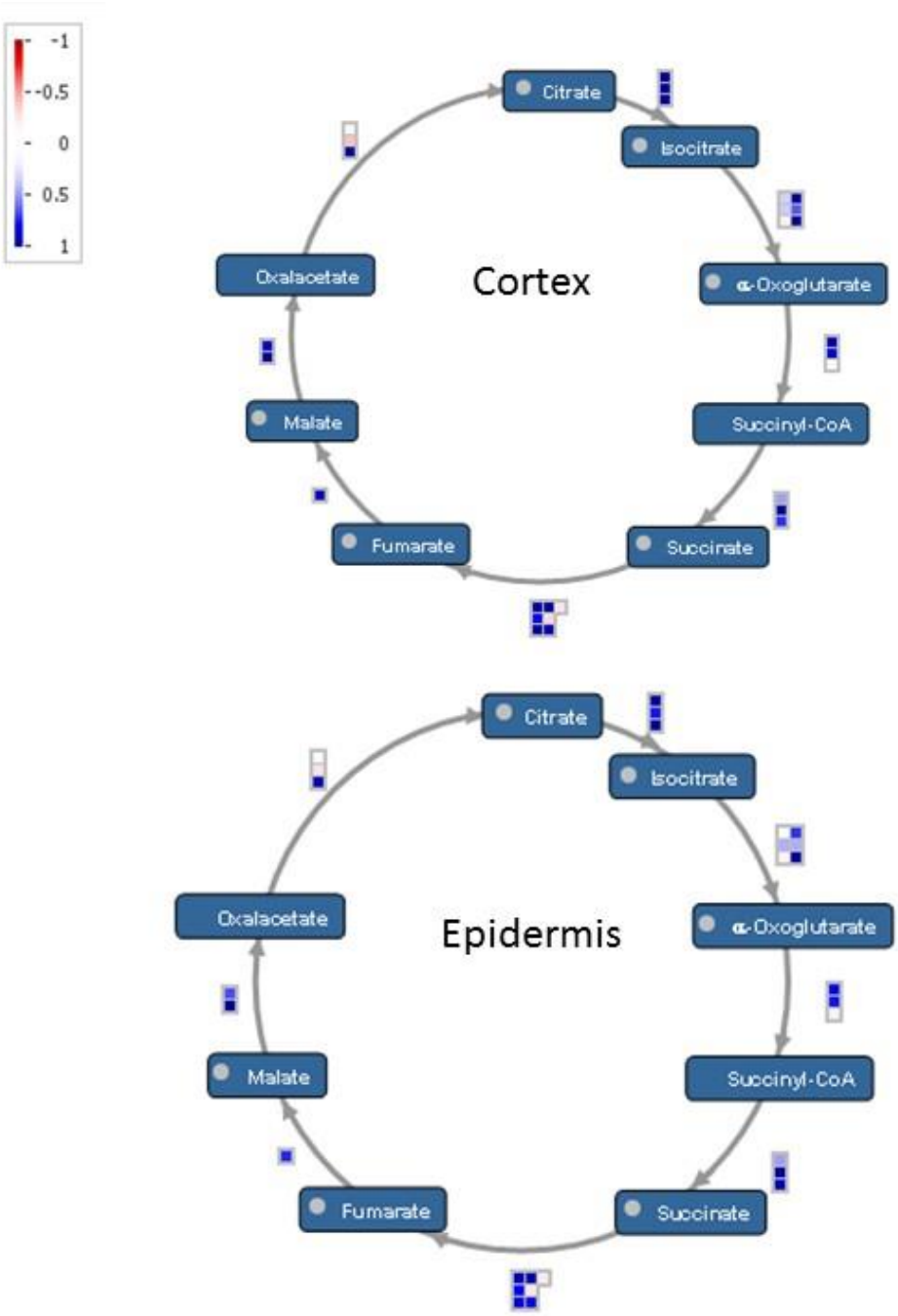
patterns emerge in genes related to processes such as mitochondrial electron transport and metabolic pathways related to lipids and cell wall synthesis. To investigate differences in the gene expression in the cell types in more detail, the data in all bins with significant variation was collated from each cell type. This comparison showed that the largest number of bins and genes showing variation in these datasets were those that varied significantly across all cell types included in the analysis (Figure 54). 115 bins containing 27220 genes were in this group. The bins contain a different number of genes related to a specific metabolic process and in some cases genes may be grouped to more than one bin. The second largest group was genes that showed significant variation in a sub-population that consisted of two or more cell types. This group consisted of 105 bins with 10087 genes. In addition, there were 57 bins with 2858 genes where significant variation in a single cell type was indicated by the p-values. Figure 55 shows the distribution of the significant differences between cell types.

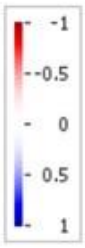
A further investigation of the types of metabolic processes showing variation based on Wilcoxon rank sum test revealed that the genes different across cell types were related to such processes as primary metabolism, protein synthesis, targeting and degradation, DNA and RNA synthesis, amino acid metabolism, signalling, transport, hormone metabolism, stress and cell wall and lipid metabolism. In some cases other specific processes were affected in a sub-set of cell types. These include phenylpropanoid and polyamine pathways. The overall analysis of the gene expression data suggests that there are cell type specific differences in gene expression linked to metabolism between these cell types: in columella cells expression of groups of genes associated with starch degradation are over represented, whereas in endodermis expression of suites of genes associated with multiple metabolic processes are significantly different in comparison to the other cell types examined. An additional significant pattern emerging from the data is the over expression of genes in pathways related to phenylpropanoids in cortex stele and endodermis cells, but not in the other cell types. Accumulation of phenylpropanoids in cortex cells has been discussed previously (Hemm et al., 2004) and demonstrated by measuring accumulation of flavonol glycosides in cortex cells (Rogers et al.,

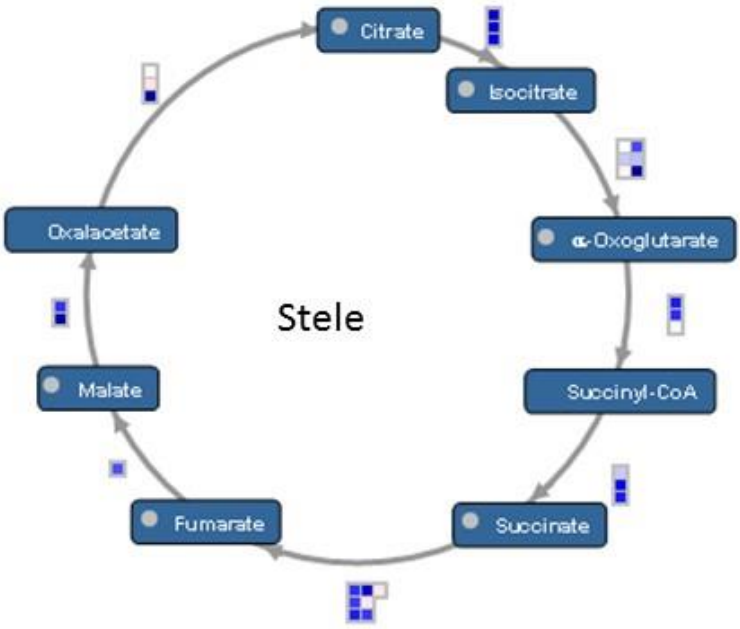
2012). Rogers and co-authors also suggest the enrichment of genes related to glycosyltransferases in these cells could be linked to the glycosylation of flavonoids.

Figure 51: **Gene expression in the TCA cycle of cells from the cortex, epidermis, columella, endodermis and stele.** Publicly available normalised microarray data for the four cell types was mapped against whole root data using the Arabidopsis mapping based on TAIR 9 genome release using MapMan software. Analysis was focused on the heat map of TCA cycle gene expression. Gene expression is presented as a false colour heat map (scale shown).




-1
-0.5
0
0.5
1
Citrate
Isocitrate
α-Oxoglutarate
Succinyl-CoA
Succinate
Fumarate
Malate
Oxalacetate
Cortex
Citrate
Isocitrate
α-Oxoglutarate
Succinyl-CoA
Succinate
Fumarate
Malate
Oxalacetate
Epidermis


-1
-0.5
0
0.5
1


Citrate
Isocitrate
α-Oxoglutarate
Succinyl-CoA
Succinate
Fumarate
Malate
Oxalacetate
Stele

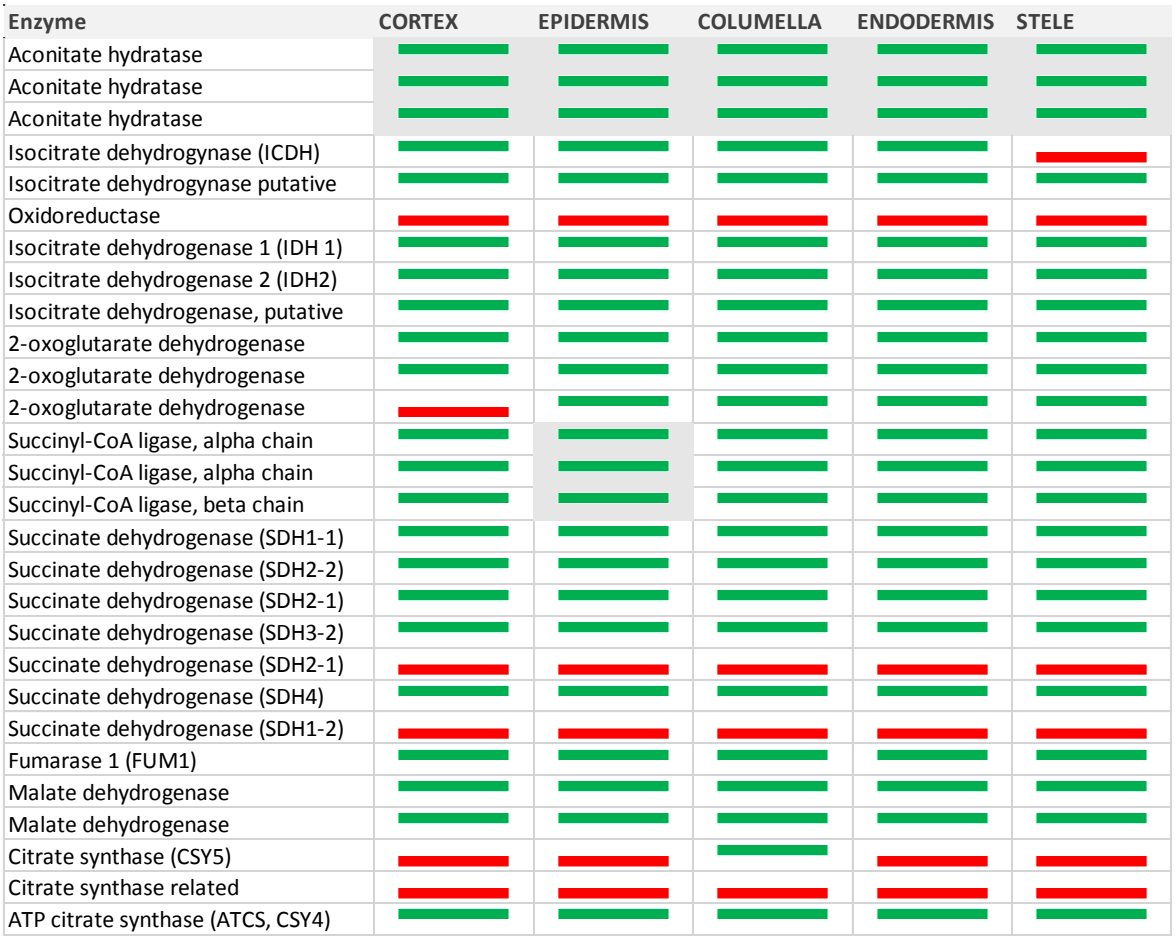
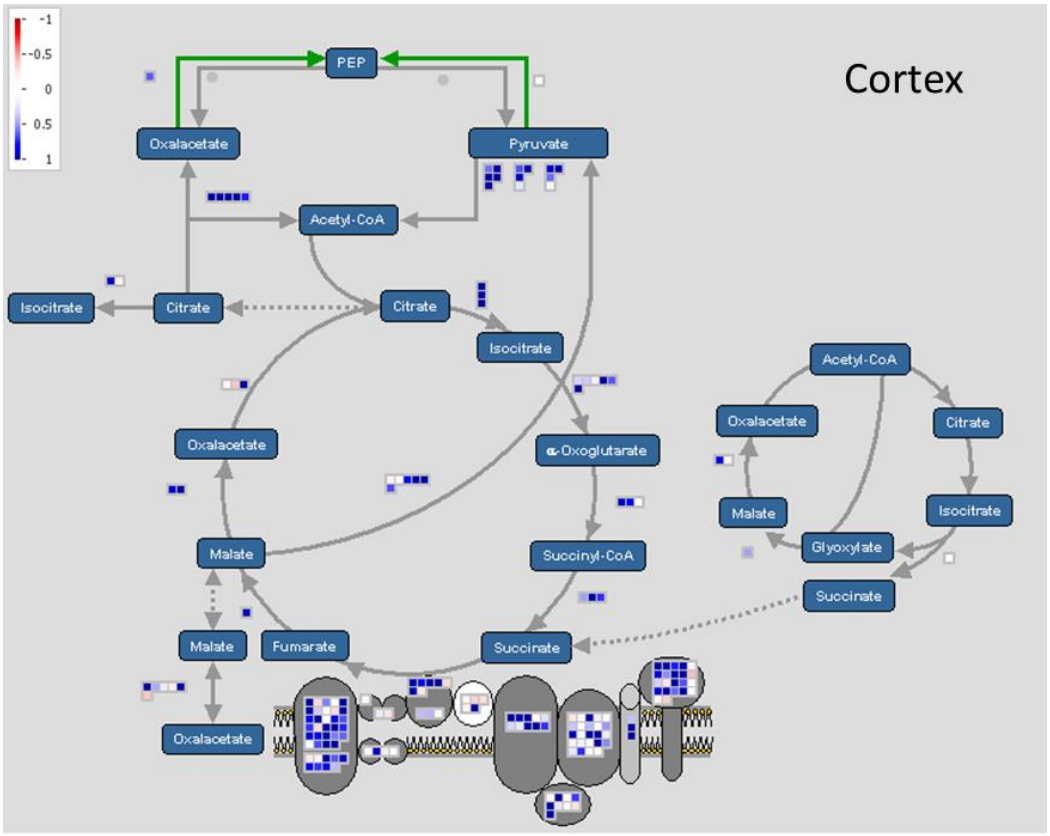
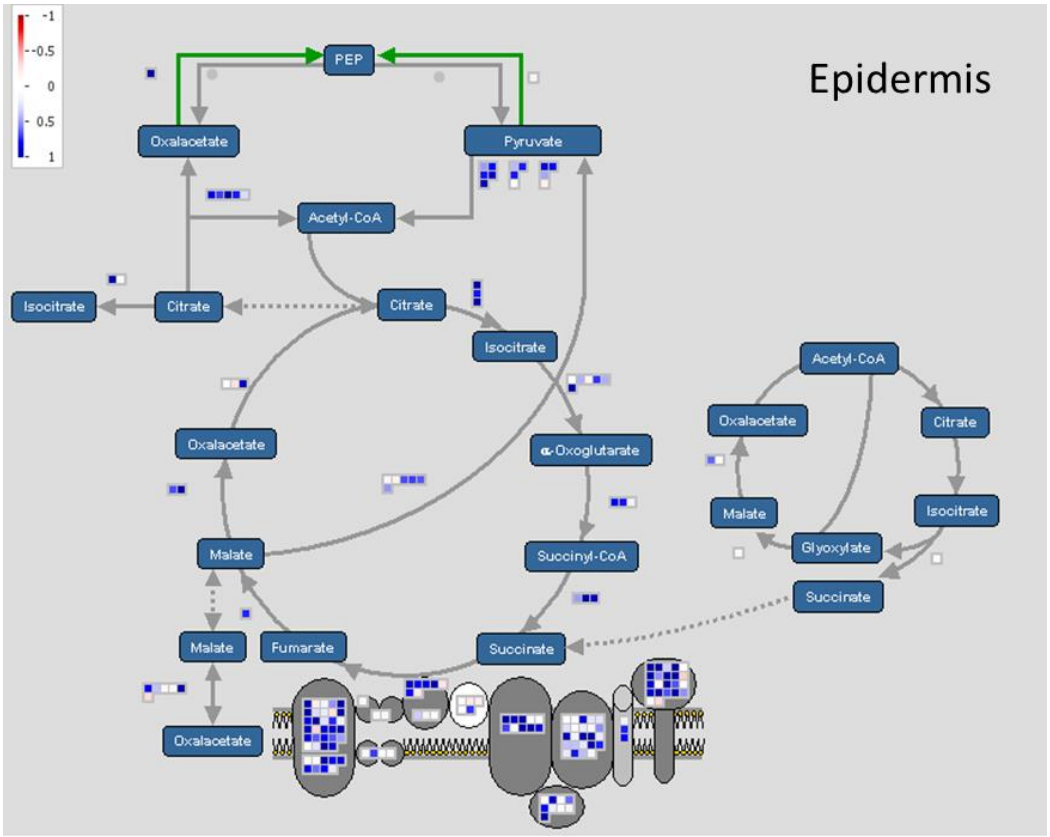
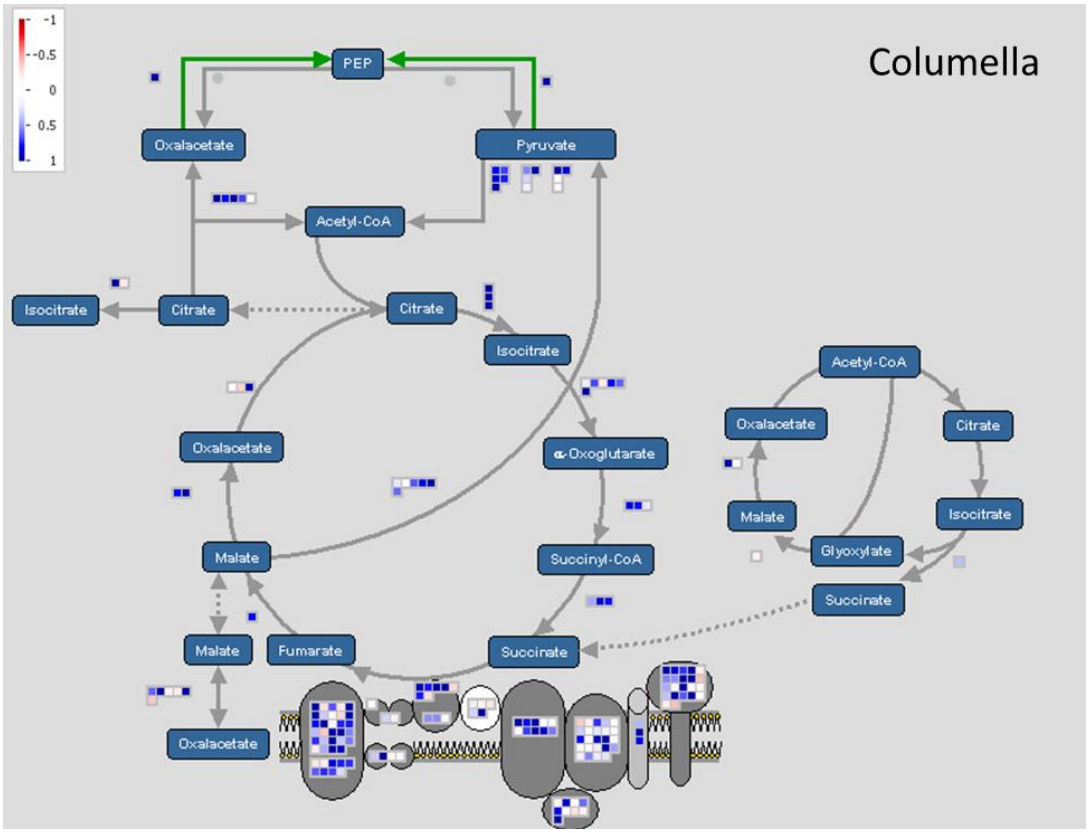
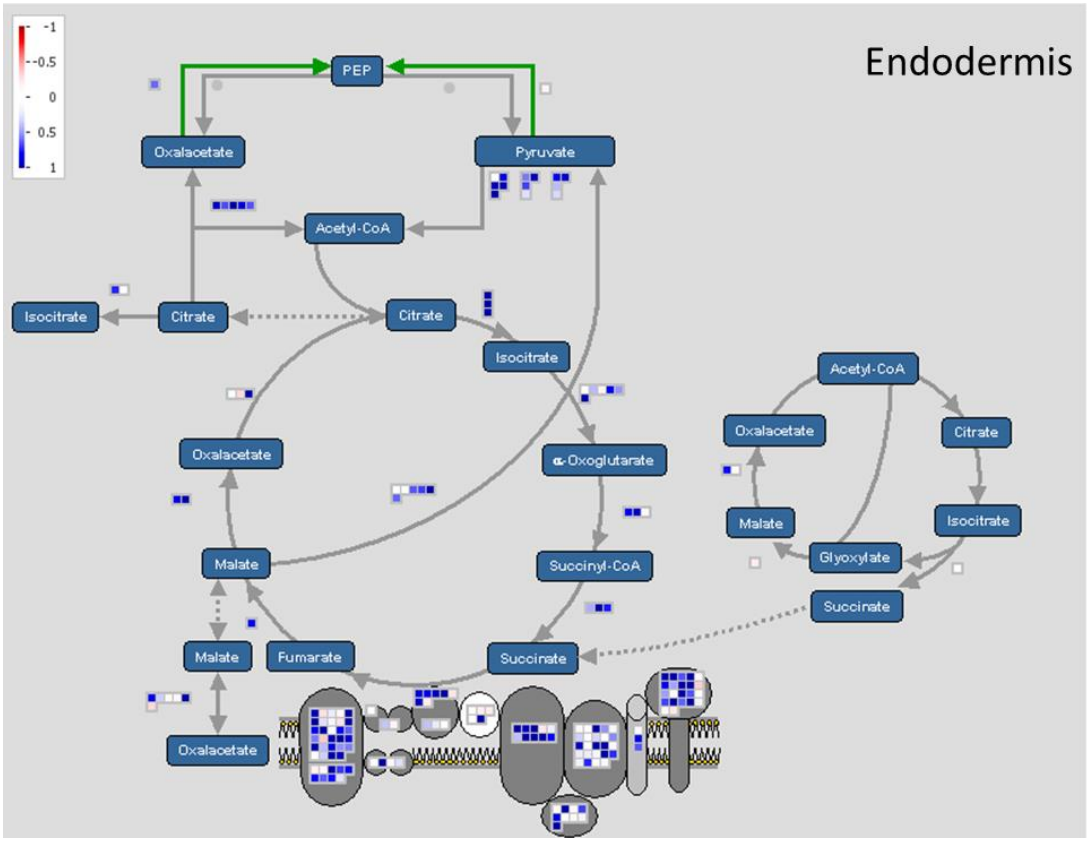


Figure 52: **Comparison of gene expression in the TCA cycle in cortex, epidermis, columella, stele and endodermis cells.** Red bars indicate lower levels of expression to wildtype data and green bars indicate up regulation. Changes with grey background were found statistically significant against the rest of the distribution using Wilcoxon rank sum test and Benjamini-Hochberg correction.


Cortex
PEP
Oxalacetate
Pyruvate
Acetyl-CoA
Isocitrate
Citrate
Citrate
Isocitrate
Acetyl-CoA
Oxalacetate
Citrate
Oxalacetate
α-Oxoglutarate
Malate
Isocitrate
Malate
Succinyl-CoA
Glyoxylate
Succinate
Malate
Fumarate
Succinate
Oxalacetate


Epidermis
PEP
Oxalacetate
Pyruvate
Acetyl-CoA
Isocitrate
Citrate
Citrate
Isocitrate
Acetyl-CoA
Oxalacetate
Citrate
Oxalacetate
α-Oxoglutarate
Malate
Isocitrate
Malate
Succinyl-CoA
Glyoxylate
Succinate
Malate
Fumarate
Succinate
Oxalacetate


Columella
-1
-0.5
0
0.5
1
PEP
Oxalacetate
Pyruvate
Acetyl-CoA
Isocitrate
Citrate
Citrate
Isocitrate
Oxalacetate
α-Oxoglutarate
Malate
Succinyl-CoA
Malate
Fumarate
Succinate
Oxalacetate
Acetyl-CoA
Oxalacetate
Citrate
Malate
Isocitrate
Glyoxylate
Succinate


Endodermis
-1
-0.5
0
0.5
1
PEP
Oxalacetate
Pyruvate
Acetyl-CoA
Isocitrate
Citrate
Citrate
Isocitrate
Oxalacetate
α-Oxoglutarate
Malate
Succinyl-CoA
Malate
Fumarate
Succinate
Oxalacetate
Acetyl-CoA
Oxalacetate
Citrate
Malate
Isocitrate
Glyoxylate
Succinate

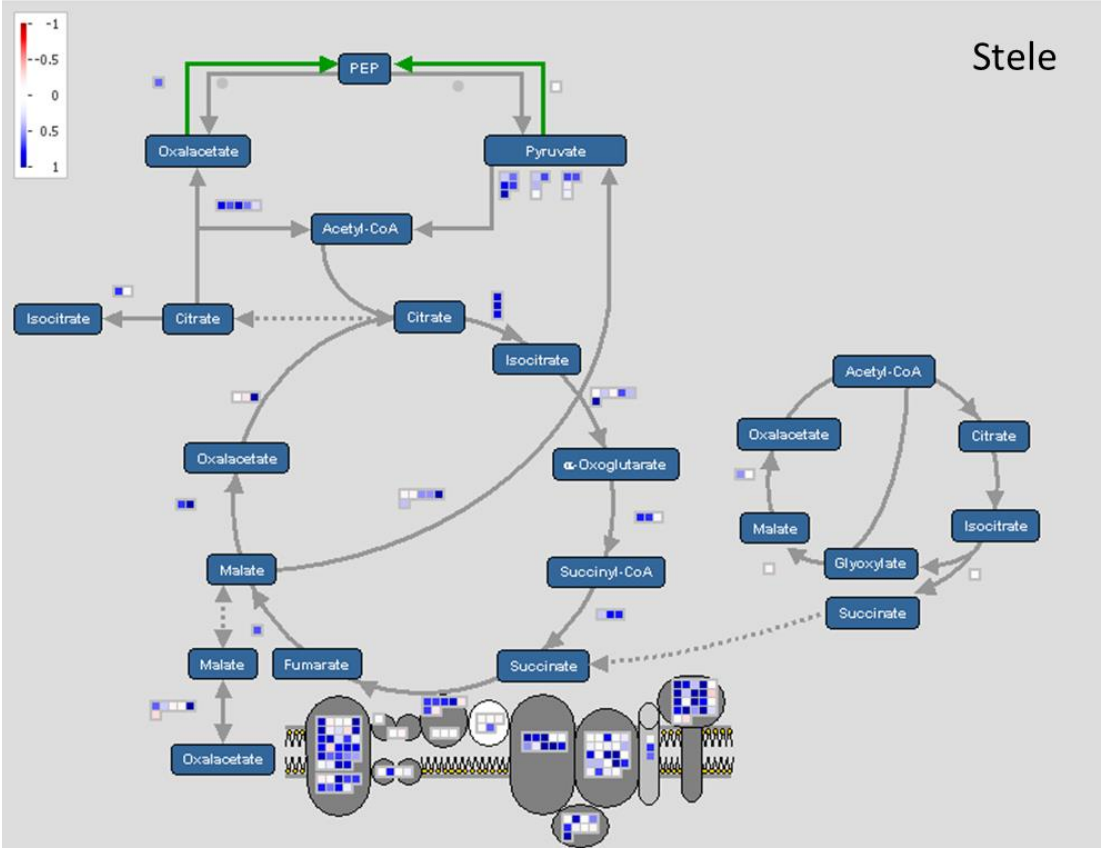
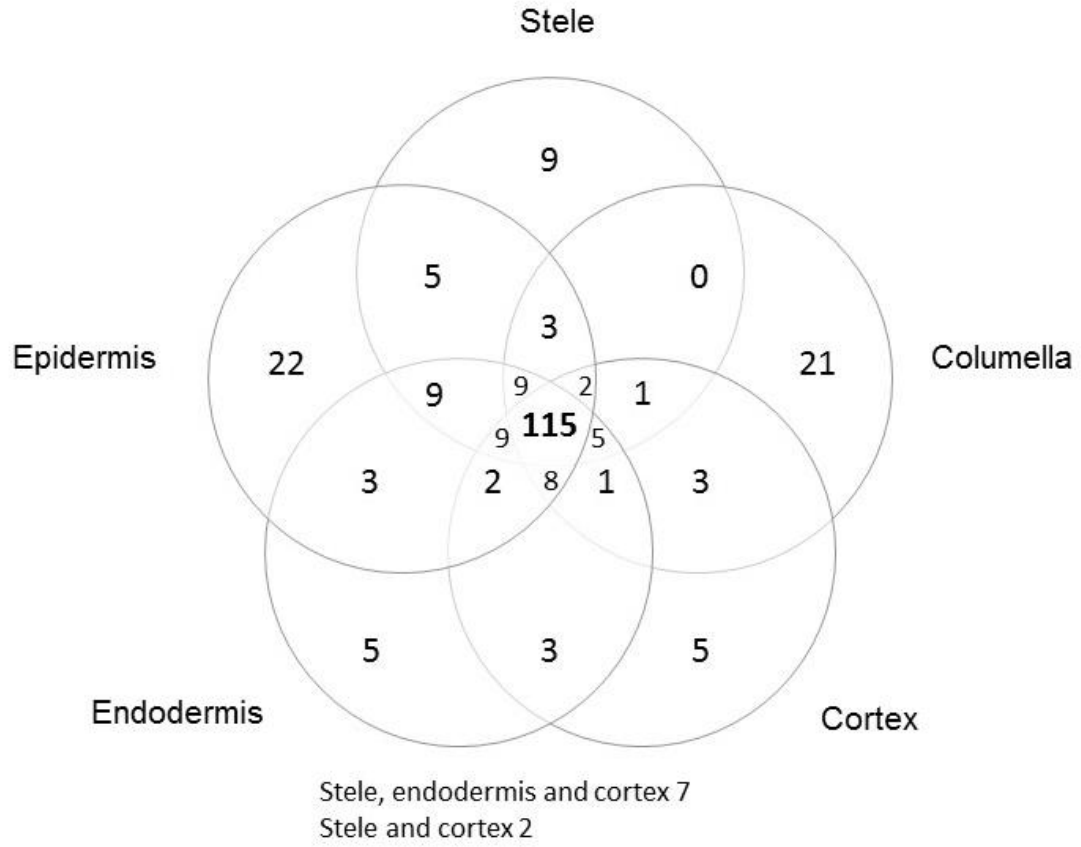


Figure 53: **Heat map of variation in expression of genes corresponding to steps in the central metabolism for cortex, epidermis, columella, endodermis and stele.** Mapping of the pathways in central metabolism against the Arabidopsis genome (TAIR 9) was done using MapMan software. The false colour heat map indicates differences to whole root data for each gene and significance of the variation was separately assessed.



Bins with variation in all cell types 115 (27220 genes)
Bins with variation in a sub-set 105 (10087 genes)
Bins with variation in one cell type 62 (2858 genes)

Figure 54: **Distribution of variation from whole root data in gene expression of specific cell types.** Data for each of the cell types was analysed separately against whole root data and bins where Wilcoxon rank sum test corrected with Benjamini-Hochberg procedure resulted in p-values below the threshold for significance (cut off at 0.05). These bins were then compared across cell types and grouped based on the combination of cell types between which significant differences variation were detected. In addition, the number of genes constituting the bins in the main groups (variation in all cell types, variation in a sub-set of cells and variation in all cell types) was collated.

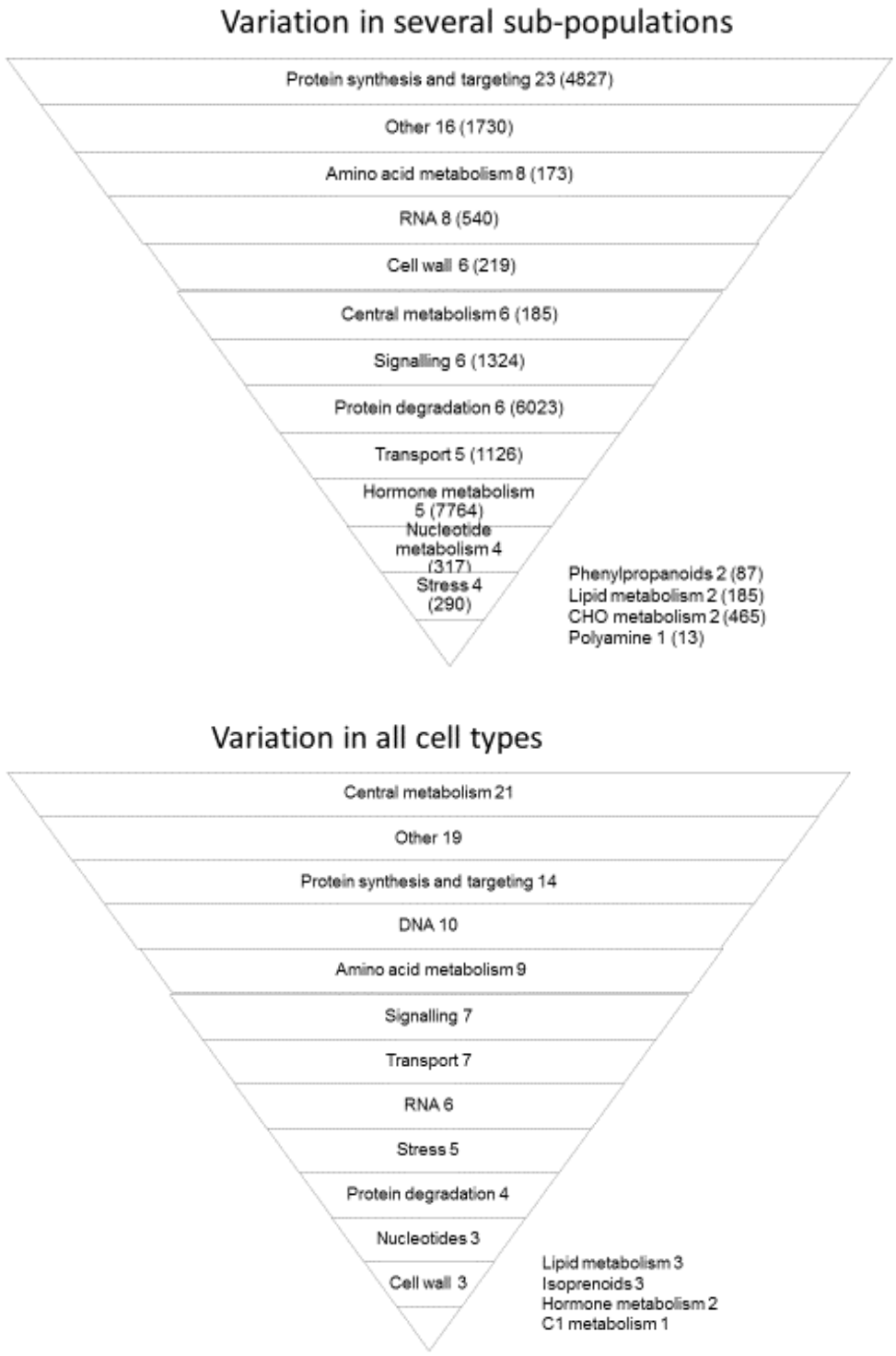


Figure 55: **List of metabolic reactions with significantly different overall gene expression pattern in all of the cell types or a sub-population of cell types.** Significance was estimated from p-values based on Wilcoxon rank sum test to identify bins with significantly different variation in the expression taking into account variation in expression of all the genes in the bin compared to the distribution in the whole dataset. Values for endodermis, cortex, epidermis, columella and stele were derived from the corresponding datasets for each of the cell populations.

7.3 Discussion

The main obstacle for cell type specific metabolic flux analysis using the reporter protein strategy is the small amount of protein that can be purified from systems such as the different cell types of the Arabidopsis root. Reliable measurement of mass isotopomer distributions remains the main challenge, but significant progress has been made in understanding the different factors that must be taken into account to achieve this and in defining the options for analysis with different analytical techniques.

In Chapter 4 modification of the standard protocols allowed the detection of mass isotopomer distributions from small quantities of protein with a standard GC-MS system, but these modifications were not sufficient for reliable detection of fragments from the cell type specific GFP samples purified from four different cell types in the Arabidopsis root. Significant variation was detected in the fractional abundances of ^{13}C enrichment for the amino acid peaks derived from GFP with cell type specific expression. The large variation detected when total ion counts are too low highlights the danger that fractional abundances of ^{13}C within the anticipated range may arise purely by chance. The variation in the ^{13}C enrichment at the lower ranges of total ion counts makes it impossible to use the data for steady state stable isotope labelling experiments where values are expected to vary due to biological differences.

In order to increase the levels of amino acids and to achieve consistent reliable detection of the mass isotopomer distributions in amino acid fragments with the standard system, the amount of biological material used could be increased. Increase in material can be achieved by increasing the number of cultures used. This approach has the disadvantages of increased costs and workload. Equally, inlets capable of large volume injection can be used to enable a significantly larger volume of protein hydrolysate to be analysed.

Alternately, a more sensitive mass spectrometer could be used to increase the detection of fragments from samples that contain low levels of contamination. If the samples do not contain high concentrations of contaminants, detection of amino acids from low concentration samples does not

require additional specificity in addition to sensitivity. Here the significant increase in levels of amino acid fragments detected using a more sensitive system seems to have had a direct effect on the ^{13}C enrichment detected from alanine from the GFP purified from the pCortex line. The use of selective methods, such as SIM and LC-MS/MS could have advantages when the complex matrix resulting both from chemical and biological contamination cannot be reduced to a minimum.

Reliable detection of mass isotopomer distributions from different cell types can most likely be achieved through modification to several aspects of the standard workflow for metabolic flux analysis. This improvement will involve a trade-off between sensitivity of detection, which may not be widely available, and the amount of protein that can be hydrolysed and injected into the analysis, with the downside of increase in cost and workload per sample. In both cases, more stringent criteria for assessing the reliability of measurement data is required to ensure the same quality of data as in standard experiments in which high amount of protein can normally be obtained from the biological material. The work described here highlights the need for implementing new workflows specific for analysis of smaller cell populations.

Cell types in the Arabidopsis roots are known to differ to the extent that it has been possible to identify single cells based on expression profiles (Efroni et al., 2015). The number of observations that suggest there may be cell type specific differences that are directly related to metabolism is also increasing. The significance of the variation seen in expression of genes related to key metabolic processes to actual metabolic fluxes is difficult to judge based on gene expression data alone (Schwender et al., 2014). Enzyme redundancy may mean some differences in gene expression patterns have no direct observable effect in central metabolism and on the other hand changes might occur through post translational modification or as a result of substrate availability without any significant change in the transcriptome (Chubukov et al., 2013; Daran-Lapujade et al., 2007). This demonstrates the need to investigate cell type specific metabolism further by combining information from different layers of the metabolic network.

The significant changes observed here in gene expression in the phenylpropanoid pathways in different cell types from Arabidopsis roots have been previously described (Birnbaum et al., 2003; Brady et al., 2007; Hemm et al., 2004) and quantified (Rogers et al., 2012). They may relate to the observations that surrounding cells support cell wall lignification of xylem cells (Birnbaum et al., 2003; Smith et al., 2013). The discovery by Smith and group that cell walls of xylem cells are additionally lignified after cell death by use of extracellular components raises interesting questions about the metabolic activity of individual cell types. Metabolic processes that are not confined to one cell type may affect the reliability of some of these metabolites as reporters on flux within cells and especially output fluxes quantified based on biomass components. Further analysis of gene expression data together with other data on the metabolism of the cells, including metabolic fluxes, can be an invaluable tool to recognise pitfalls in defining these type of processes taking place in metabolic networks of specific cell types.

Another area of research that can help in investigating cell type specific metabolism is measuring metabolite levels. Moussaief and colleagues carried out a metabolomics experiment on five specific cell type populations using UPLC-qTOF-MS on protoplasts sorted by FACS followed by further identification of some of the detected masses by MS/MS fragmentation (Moussaieff et al., 2013). Similar to the work on the gene expression patterns, the group found differences in levels of phenylpropanoids, but also in glucosinolates and dipeptides. For glucosinolates, higher abundances were detected in cortex and columella cells with the aliphatic class more abundant in cortex and indole glucosinolates more abundant in cells from the columella. Levels of glucosinolates were lower in endodermis and stele. Dipeptides, on the other hand, were detected at higher levels from all cell types and especially from endodermis and epidermis cells. These changes were not related to significant changes in transcript levels. In addition, the authors found that the abundances of several metabolites were significantly higher in isolated cells compared to the whole root and suggest analysis at the level of specific cell populations may elucidate differences, which cannot be measured from the mixture of information on cells in the whole root data.

Another very recent example of metabolomics analysis of the cell types in Arabidopsis roots also supports the hypothesis that the cells have differences in their central carbon metabolism. Petersson and colleagues analysed FACS isolated protoplasts obtained from the Arabidopsis root with Agilent 7890A gas chromatography system and a high throughput TOF-MS (Petersson et al., 2015). They found the protoplasts from the endodermis and cortex contained in general higher quantities of long chain carbohydrates, which were not found in the whole root samples, potentially due to low concentrations. Higher abundances of metabolites such as asparagine, α-ketoglutarate, glutamine, glutarate, malate, citrate and phosphoric acid were also correlated with the protoplasts. In conclusion, the metabolite profiles of cortex and endodermis cells were significantly different from reference cells when subjected to principal component analysis (PCA). These results appear to agree well with the gene expression data analysed here, however, the analysis by Petersson and group is limited to a small number of cell types.

An unusual characteristic of the gene expression data obtained from protoplasts of cell populations is the number of genes that are enriched across cell types. It is not immediately clear why whole root data would be different from all of the cell types. Conclusions from published studies on cell type specific metabolism such as enrichment of phenylpropanoid pathways agree with the analysis presented here. However, the large number of genes and bins related to key pathways in central metabolism that are significantly different across the cell populations has not been discussed in detail.

Gene clustering approaches have yielded a range of important data on cell type specific gene expression and confirmed earlier published work on cell type specific processes (Birnbaum et al., 2003; Dinneny et al., 2008). As part of this work, the effect of protoplasting has been investigated: Birnbaum and group used a control population of Arabidopsis cells obtained from the same enzymatic digestion of the cell wall as the cells that were sorted with FACS and collected the control cells in increments over one hour to simulate the conditions used for sorting. In this control test over 300 genes were found to be affected by the cell isolation alone, compared to the control group of cells. Similar controls were applied by Petersson and group in comparing a reference group of cells

to the protoplasts that were used in analysis of metabolite levels. Moussaief and group on the other hand used data on untreated whole root compared against protoplasts derived from the same plant material and sorted by FACS without a GFP gate to subtract any changes that were due to protoplasting from their metabolite data: all data where higher intensity correlated with protoplasts was removed from analysis (Moussaieff et al., 2013). The last experiment accounts for the effect of enzymatic digestion and changes that occur due to delay from preparation of the samples before further processing.

As the authors state, even this type of procedure cannot completely exclude the possibility of effects from protoplasting (Moussaieff et al., 2013). For example, Fesenko and group found that cells from the moss *Physcomitrella patens* experience a burst of digestion of proteins into peptides as a result of protoplasting (Fesenko et al., 2015). The peptidome of the protoplasts contained a large number of peptides derived from proteins in the chloroplast. Fesenko and colleagues also confirmed that tissue treated with lower concentrations of the digestion enzymes used for protoplasting resulted in increase of dipeptides in the tissue, even when protoplasts were not formed. When the group analysed the gene expression in these cells, they found genes directly related to protein degradation were enriched in the protoplasts together with several classes of genes related to stress responses. In addition to any direct effects, metabolic processes could be affected in the cells within a much shorter time frame than is required for the preparation of samples, digest and FACS. Especially in the case of metabolites rather than RNA, turnover times may be seconds rather than hours (Fernie et al., 2011).

Nevertheless, differences in levels of dipeptides between cell types suggest there are underlying differences in processes that are directly linked to central metabolism in these cells. Similarly there seem to be cell type specific differences in expression of genes encoding for enzymes with a role in the TCA cycle and those directly linked to amino acid metabolism. Therefore significant differences in central metabolism of these cell types are difficult to rule out without further investigation into metabolic activity of the cells. Such differences could be observed by quantifying the metabolic activity at the level of fluxes. The reporter protein strategy has the additional advantage that cells

can be directly frozen in liquid nitrogen prior to extraction steps, which quenches metabolic processes before the tissue is disrupted. The work presented in this chapter shows measurements of isotopomer data from this system could be feasible with the right analytical techniques and a modified workflow. Together these approaches can in the future provide detailed systems level understanding of the metabolism in the different cell types in Arabidopsis roots.

8 General Discussion

8.1 Summary of results

The validation of GFP as a reporter protein for metabolic activity in the Arabidopsis roots described in this work was achieved by 1) developing a robust experimental and analytical strategy for detection of mass isotopomer distributions from small amounts of amino acids and 2) validating the use of the reporter protein in the biological system. The first part of this work focused on the development of a highly specific purification using GFP nanotrap, the introduction of purification steps to address biological and chemical contamination and on establishing the protocols for sample preparation for detection of small amounts of amino acids derived from the reporter protein with GC-MS. These protocols developed in Chapter 4 were applied in the second part, where the suitability of GFP as a reporter in the Arabidopsis root was evaluated.

This work involved three main approaches. The first demonstrated the ability of GFP to report on the isotopic labelling normally measured from the total protein fraction, showing that the labelling information could be used for steady state stable isotope MFA. The second investigated the fractional abundances of ^{13}C in amino acids derived from protein extracted from leaves and roots of seedlings grown in 20% [$^{13}C_6$]-glucose and suggested there was a negligible contribution from CO_2 assimilation through photosynthesis to root and leaf metabolism under the experimental conditions. The third approach compared fractional abundances of ^{13}C in amino acids derived from the total protein fractions from wildtype seedlings grown in 20% [$^{13}C_6$]-glucose to those derived from seedlings with GFP over expression. The over expression had minimal effect on mass isotopomer distributions detected from amino acids, suggesting the levels of expression in this system do not perturb central metabolism. Together the results from Chapter 5 confirm the suitability of GFP as a reporter protein in the Arabidopsis root and provide important information about the system. The detected levels of fractional abundances of ^{13}C in amino acids derived from all protein samples that had been extracted from seedlings grown in 20% [$^{13}C_6$]-glucose were at approximately 20%. The equal ^{13}C enrichment between the amino acids and the glucose substrate suggests the amino acids in protein reached an isotopic steady state within the 14 days of growth.

In chapter 7 the protocols that had been developed were applied to an investigation of specific cell types in the Arabidopsis root. The detection of reliable mass isotopomer distributions from the GFP purified from the cells proved difficult due to the lower levels of protein that was extracted, but the work established the framework for detection of cell type specific metabolic activity in four different cell types in the Arabidopsis root and the approaches that can be taken are outlined. In addition, the gene expression patterns in five of the cell types were analysed. This analysis yielded results that show the possibility of cell type specific differences in central carbon metabolism cannot be ruled out without further investigation into the metabolic activity of the cells.

In Chapter 6 surprisingly different fractional abundances of ^{13}C were measured from isolated root hair cells in comparison to the whole root total protein as part of developing a model system for cell type specific metabolism. The root hair total protein fraction suggested a significantly lower fractional abundance of ^{13}C in amino acids derived from protein compared to the whole root total protein fraction. Free amino acids on the other hand had reached approximately 20% ^{13}C enrichment in root hairs. Thus it seems root hair cells may not have reached isotopic steady state in 14 days, despite the average data on the whole root suggesting isotopic labelling equal to the glucose substrate.

There may be differences between the wildtype root hairs and root hairs isolated from the over expression line: the ^{13}C enrichment in amino acids derived from the RSL4 line was indistinguishable from amino acids derived from wildtype whole root total protein fraction. The differences in amino acid labelling between the two types of root hairs are not readily explained by any analytical or experimental factor and they may indicate the root hair cells take a longer time to reach isotopic steady state than the cells in the root on average. If proven correct, this hypothesis could have implications for metabolic models of root hair metabolism. Specifically, the models would have to take into account an unlabelled protein pool present at the start of the experiment (Kruger et al., 2012; Lonien and Schwender, 2009; Williams et al., 2008).

Chapter 3 focused on elucidating the requirements for modelling heterogeneous tissue which contains separate metabolite pools by studying whether average data reports on the aggregate flux distribution. Because of the nonlinear relationship between isotopomer data and the deduced fluxes, aggregate flux maps were found to be unreliable representations of the actual fluxes in the system. An exception to this would be, if all the fluxes in the cells contributing to the aggregate flux are highly similar. The future modelling approach for differentiated tissues therefore depends on how different the flux phenotypes of different cell types are.

In conclusion, applying the reporter protein strategy described in this work provides one way of investigating the metabolic activity that needs to be understood better in order to develop MFA further. The basis for such investigations in the Arabidopsis root have been outlined for four cell types and separated root hairs are suggested as an additional model system for cell type specific metabolism, with the additional benefit of linking cell type specific MFA with the wider system level analysis of the Arabidopsis root.

8.2 Current mass spectrometry techniques have the capabilities required for cell type specific MFA with the reporter protein strategy

Detection of isotopic labelling from small amounts of amino acids derived from a biological sample is a complex analytical challenge. Indications are that the instrument capable of routine analysis of these samples would most likely combine high sensitivity with large injection volumes. The levels of amino acids that were detected using such systems in Chapter 4 demonstrate on a general level that mass spectrometry instruments can detect high total ion counts from samples that contain very low concentrations of amino acids, as long as the samples are clean of major contaminants. Including selective modes such as SIM in the protocols may also prove successful.

The major limitation for these methods does not seem to be the sensitivity of the analytical instruments alone, but the combination of factors that need to be addressed by a specific analytical strategy. However, this is not specific to MFA, but often faced by other fields relying on mass spectrometric detection of proteins and metabolites. Cell type specific analysis will benefit from new

approaches and solutions with higher sensitivity. Some of the solutions for these issues may arise from a wider awareness of the existing analytical techniques, such as large volume injection.

Experiments with new analytical techniques require validation, as demonstrated by published work on measuring isotope labelling from peptide fragments. When fragments were produced through collision induced dissociation, isotopic composition seemed to be disproportionately presented in the analysed fraction compared to the precursor molecules (Allen et al., 2014a). The authors concluded the use of the data for flux analysis would be limited by the deviations from predicted distributions of isotopes. Work to assess the disparity of different approaches for measuring isotope labelling in parent ions and fragmented ions as well as instrument and protocol specific effects could help in future work. A recent multi-instrument inter-laboratory metabolomics study with conditions that had not been standardised found that the spectral information extracted from data of the same samples run on several LC-MS and NMR instruments using different methods showed high convergence (Martin et al., 2015). Similar work on confirming how the technique and instrument specific aspects affect measurements in case of isotopic labelling may provide further insight on which techniques are most promising for cell type specific MFA outside of traditional GC-MS analysis.

In addition, experiments relying on quantification of isotope labelling from small amounts of analytes will benefit from validation against criteria such as limit of blank, limit of detection and limit of quantitation (Armbruster and Pry, 2008). Introducing such controls routinely used across other fields of mass spectrometry to analyse trace levels of metabolites may help in developing robust strategies for cell type specific analysis.

8.3 Future work

Directions for future research are outlined by the work described in Chapter 6 and Chapter 7. In these chapters, work was started on developing a model system for cell type specific metabolism and investigating cell type specific metabolic activity in the Arabidopsis root, respectively. In Chapter 7, steps that must be taken in order to detect mass isotopomer distributions from the cell type specific GFP were described as dependent on the analytical instrument and type of samples used. The work

described in this thesis can be used as the basis for experiments investigating cell type specific metabolism in the Arabidopsis root. The limits of the analytical techniques have not yet been reached and further work in this biological system can yield important information on the potential differences between cell types.

Similarly, developing a model system using isolated root hair cells will not only be important in order to develop modelling approaches for MFA, but can provide biologically interesting information on root hair cells. This information can potentially be linked with the other systems analysis approaches that are being developed (Libault et al., 2010). A strategy to study metabolic activity of specific cell types is required to further the understanding of this layer of the metabolic network, but additionally to guide development of MFA. The work described in this thesis provides one such strategy.

8.4 Conclusions

The primary aim of this thesis was to validate GFP as a reporter protein for cell type specific metabolism in Arabidopsis roots and to investigate the feasibility of using the reporter protein strategy in this biological system. A reporter protein strategy provides one solution for cell type specific analysis of metabolic activity, which is not possible through current standard techniques for MFA. A protocol has been developed based on highly specific immunopurification of GFP and modifications made to traditional methods used to analyse stable isotope labelling in amino acids derived from protein. The use of this protocol in hydroponically grown seedlings of Arabidopsis lines with cell type specific gene expression has been validated.

Additionally, this work has demonstrated the need to carefully assess the quality of mass isotopomer data obtained from small amounts of derivatised amino acids analysed with GC-MS: data on low concentration samples does not converge to a consistent measurement value when run with the same instrument using the same protocols and methods. However, at higher concentrations data obtained from the same samples reports well on the expected ^{13}C enrichment provided in the substrate. A higher concentration can be achieved either through changes to the protocols or changes to the

instrument used to analyse the samples. The issue may therefore be partly resolved by use of sensitive instruments and techniques such as large volume injection.

Further to this, findings in this thesis demonstrate the need to explore cell type specific metabolism further. Significant variation in expression in genes in the central metabolism between the cell types suggests metabolic activity in these cell should not be assumed to be the same for all metabolic processes. Increased understanding of metabolism in multicellular systems will guide development of approaches for cell type specific metabolic models. It is needed, since data obtained from a mixture of cells in these systems may not provide an accurate way of modelling flux distributions.

Finally, root hairs have been investigated in this thesis as a potential model system for cell type specific metabolism. These cells can be easily isolated from the root and therefore it becomes possible to obtain cell type specific data on soluble metabolites and cell wall components normally analysed in MFA experiments. Work in this system should in the future provide an expanded view on best approaches for modelling cell type specific metabolism with data obtained through the reporter protein strategy. Whether metabolic models need to be constructed at the level of specific cell types due to significant differences in central metabolism in these cell types remains to be investigated.

Appendix list:

Appendix 11: **Mass spectra of amino acids derived from immunopurified GFP from three cultures of 30-50 seedlings hydrolysed and derivatised with a reduced total volume of derivatisation reagents (20 µl).**

Appendix 12: **Comparison of total ion counts and fractional abundances for individual amino acid fragments detected in tBDMS derivatised protein hydrolysate.**

Appendix 13: **Examples of fragments not present at consistently high enough levels for analysis and not included in the data collated in Table 6.**

Appendix 14: **Mass spectra of hydrolysed GFP immunopurified from a single culture of 30-50 seedlings with constitutive GFP expression and derivatised in a reduced volume of reagents.**

Appendix 15: **Comparison of mass spectra of amino acid fragments detected using Agilent 5975 and 7200 systems.**

Appendix 16: **Comparison of average ^{13}C percentage derived from amino acid fragments detected from purified GFP samples and the same fragments from the total protein fraction.**

Appendix 17: **Comparison of fractional abundances of ^{13}C in amino acid fragments detected from the total protein fractions from root and shoot tissue in Arabidopsis seedlings.**

Appendix 18: **Comparison of fractional abundance of ^{13}C detected from total protein fractions from wildtype seedlings and seedlings with constitutive GFP expression**.

Appendix 19: **Comparison of the mass isotopomer distribution of amino acids derived from GFP and total protein in E. coli. Reanalysis of data presented in Shaikh et al. (2008)**

Appendix 20**: Mass isotopomer distributions in amino acid fragments derived from hydrolysed total protein fraction and GFP from Arabidopsis seedlings with standard deviations.**

Appendix 21: **Statistical comparison of average ^{13}C enrichment in protein derived from isolated root hairs and root total protein fraction** (Benjamini and Hochberg, 1995)**.**

Appendix 22: **Comparison of levels of metabolites detected from exudates from the wildtype Columbia seedlings and seedlings with RSL4 over expression.**

Appendix 23: **Levels of amino acid fragments detected from GFP purified from specific cell types of Arabidopsis seedlings from 5 cultures of 30-50 seedlings combined in each sample, mass peak corresponding to fragment containing no ^{13}C substitutions.**

References

Alagesan, S., Gaudana, S.B., Sinha, A., and Wangikar, P.P. (2013). Metabolic flux analysis of Cyanothece sp. ATCC 51142 under mixotrophic conditions. Photosynth Res *118*, 191-198.

Allen, D.K., Evans, B.S., and Libourel, I.G. (2014a). Analysis of isotopic labeling in peptide fragments by tandem mass spectrometry. PLoS One *9*, e91537.

Allen, D.K., Goldford, J., Gierse, J.K., Mandy, D., Diepenbrock, C., and Libourel, I.G. (2014e). Quantification of peptide m/z distributions from ^{13}C-labeled cultures with high-resolution mass spectrometry. Anal Chem *86*, 1894-1901.

Allen, D.K., Laclair, R.W., Ohlrogge, J.B., and Shachar-Hill, Y. (2012). Isotope labelling of Rubisco subunits provides in vivo information on subcellular biosynthesis and exchange of amino acids between compartments. Plant Cell Environ *35*, 1232-1244.

Allen, D.K., Libourel, I.G., and Shachar-Hill, Y. (2009a). Metabolic flux analysis in plants: coping with complexity. Plant Cell Environ *32*, 1241-1257.

Allen, D.K., Ohlrogge, J.B., and Shachar-Hill, Y. (2009b). The role of light in soybean seed filling metabolism. Plant J *58*, 220-234.

Allen, D.K., and Young, J.D. (2013). Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. Plant Physiol *161*, 1458-1475.

Allison, D.G., and Sattenstall, M.A. (2007). The influence of green fluorescent protein incorporation on bacterial physiology: a note of caution. J Appl Microbiol *103*, 318-324.

Alonso, A.P., Goffman, F.D., Ohlrogge, J.B., and Shachar-Hill, Y. (2007a). Carbon conversion efficiency and central metabolic fluxes in developing sunflower (*Helianthus annuus L.*) embryos. Plant J *52*, 296-308.

Alonso, A.P., Raymond, P., Hernould, M., Rondeau-Mouro, C., de Graaf, A., Chourey, P., Lahaye, M., Shachar-Hill, Y., Rolin, D., and Dieuaide-Noubhani, M. (2007b). A metabolic flux analysis to study the role of sucrose synthase in the regulation of the carbon partitioning in central metabolism in maize root tips. Metab Eng *9*, 419-432.

Alonso, A.P., Raymond, P., Rolin, D., and Dieuaide-Noubhani, M. (2007c). Substrate cycles in the central metabolism of maize root tips under hypoxia. Phytochemistry *68*, 2222-2231.

Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins *81*, 1857-1861.

Antoniewicz, M.R., Kelleher, J.K., and Stephanopoulos, G. (2006). Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab Eng *8*, 324-337.

Antoniewicz, M.R., Kelleher, J.K., and Stephanopoulos, G. (2007). Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal Chem *79*, 7554-7559.

Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R., and Muyldermans, S. (1997). Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS Lett *414*, 521-526.

Armbruster, D.A., and Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. Clin Biochem Rev *29 Suppl 1*, S49-52.

Badri, D.V., Chaparro, J.M., Zhang, R., Shen, Q., and Vivanco, J.M. (2013). Application of natural blends of phytochemicals derived from the root exudates of Arabidopsis to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. J Biol Chem *288*, 4502-4512.

Bargmann, B.O., Vanneste, S., Krouk, G., Nawy, T., Efroni, I., Shani, E., Choe, G., Friml, J., Bergmann, D.C., Estelle, M., et al. (2013). A map of cell type-specific auxin responses. Mol Syst Biol *9*, 688.

Becker, J.D., Takeda, S., Borges, F., Dolan, L., and Feijo, J.A. (2014). Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. BMC Plant Biol *14*, 197.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289-300.

Beste, D.J., Bonde, B., Hawkins, N., Ward, J.L., Beale, M.H., Noack, S., Noh, K., Kruger, N.J., Ratcliffe, R.G., and McFadden, J. (2011). ^{13}C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation. PLoS Pathog *7*, e1002091.

Birnbaum, K., Jung, J.W., Wang, J.Y., Lambert, G.M., Hirst, J.A., Galbraith, D.W., and Benfey, P.N. (2005). Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat Methods *2*, 615-619.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science *302*, 1956-1960.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science *318*, 801-806.

Brandt, S., Kehr, J., Walz, C., Imlau, A., Willmitzer, L., and Fisahn, J. (1999). Technical Advance: A rapid method for detection of plant gene transcripts from single epidermal, mesophyll and companion cells of intact leaves. Plant J *20*, 245-250.

Brinch-Pedersen, H., Galili, G., Knudsen, S., and Holm, P.B. (1996). Engineering of the aspartate family biosynthetic pathway in barley (*Hordeum vulgare L.*) by transformation with heterologous genes encoding feed-back-insensitive aspartate kinase and dihydrodipicolinate synthase. Plant Mol Biol *32*, 611-620.

Burrell, M., Earnshaw, C., and Clench, M. (2007). Imaging Matrix Assisted Laser Desorption Ionization Mass Spectrometry: a technique to map plant metabolites within tissues at high spatial resolution. J Exp Bot *58*, 757-763.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol *26*, 1301-1308.

Cavatorta, V., Sforza, S., Mastrobuoni, G., Pieraccini, G., Francese, S., Moneti, G., Dossena, A., Pastorello, E.A., and Marchelli, R. (2009). Unambiguous characterization and tissue localization of Pru P 3 peach allergen by electrospray mass spectrometry and MALDI imaging. J Mass Spectrom *44*, 891-897.

Cha, S., Zhang, H., Ilarslan, H.I., Wurtele, E.S., Brachova, L., Nikolau, B.J., and Yeung, E.S. (2008). Direct profiling and imaging of plant metabolites in intact tissues by using colloidal graphite-assisted laser desorption ionization mass spectrometry. Plant J *55*, 348-360.

Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R.J., Jules, M., Link, H., Aymerich, S., Stelling, J., and Sauer, U. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. Mol Syst Biol *9*, 709.

Clode, P.L., Kilburn, M.R., Jones, D.L., Stockdale, E.A., Cliff, J.B., Herrmann, A.M., and Murphy, D.V. (2009). In situ mapping of nutrient uptake in the rhizosphere using nanoscale secondary ion mass spectrometry. Plant Physiol *151*, 1751-1757.

Conrath, K.E., Lauwereys, M., Galleni, M., Matagne, A., Frere, J.M., Kinne, J., Wyns, L., and Muyldermans, S. (2001). Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae. Antimicrob Agents Chemother *45*, 2807-2812.

Daran-Lapujade, P., Rossell, S., van Gulik, W.M., Luttik, M.A., de Groot, M.J., Slijper, M., Heck, A.J., Daran, J.M., de Winde, J.H., Westerhoff, H.V., Pronk, J.T., Bakker, B.M. (2007) The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc Natl Acad Sci U S A *104*, 15753-15758.

Dauner, M., and Sauer, U. (2000). GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. Biotechnol Prog *16*, 642-649.

Day, R.C., Grossniklaus, U., and Macknight, R.C. (2005). Be more specific! Laser-assisted microdissection of plant cells. Trends Plant Sci *10*, 397-406.

Diaz de la Garza, R.I., Gregory, J.F., 3rd, and Hanson, A.D. (2007). Folate biofortification of tomato fruit. Proc Natl Acad Sci U S A *104*, 4218-4222.

Dieuaide-Noubhani, M., Raffard, G., Canioni, P., Pradet, A., and Raymond, P. (1995). Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from $_{13}$C- or $_{14}$C-labeled glucose. J Biol Chem *270*, 13147-13159.

Dinneny, J.R., Long, T.A., Wang, J.Y., Jung, J.W., Mace, D., Pointer, S., Barron, C., Brady, S.M., Schiefelbein, J., and Benfey, P.N. (2008). Cell identity mediates the response of Arabidopsis roots to abiotic stress. Science *320*, 942-945.

Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., and Scheres, B. (1993). Cellular organisation of the Arabidopsis thaliana root. Development *119*, 71-84.

Eckstein, A., Zięba, P., and Gabryś, H. (2011). Sugar and light effects on the condition of the photosynthetic apparatus of *Arabidopsis thaliana* cultured in vitro. J. Plant Growth Regul *31*, 90-101.

Efroni, I., Ip, P.L., Nawy, T., Mello, A., and Birnbaum, K.D. (2015). Quantification of cell identity from single-cell gene expression profiles. Genome Biol *16*, 9.

Endo, M., Shimizu, H., Nohales, M.A., Araki, T., and Kay, S.A. (2014). Tissue-specific clocks in Arabidopsis show asymmetric coupling. Nature *515*, 419-422.

Enfissi, E.M., Barneche, F., Ahmed, I., Lichtle, C., Gerrish, C., McQuinn, R.P., Giovannoni, J.J., Lopez-Juez, E., Bowler, C., Bramley, P.M., et al. (2010). Integrative transcript and metabolite

analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. Plant Cell *22*, 1190-1215.

Espina, V., Wulfkuhle, J.D., Calvert, V.S., VanMeter, A., Zhou, W., Coukos, G., Geho, D.H., Petricoin, E.F., 3rd, and Liotta, L.A. (2006). Laser-capture microdissection. Nat Protoc *1*, 586-603.

Fernie, A.R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., Carroll, A.J., Saito, K., Fraser, P.D., and DeLuca, V. (2011). Recommendations for reporting metabolite data. Plant Cell *23*, 2477-2482.

Fernie, A.R., Roscher, A., Ratcliffe, R.G., and Kruger, N.J. (2001). Fructose 2,6-bisphosphate activates pyrophosphate: fructose-6-phosphate 1-phosphotransferase and increases triose phosphate to hexose phosphate cycling in heterotrophic cells. Planta *212*, 250-263.

Fesenko, I.A., Arapidi, G.P., Skripnikov, A.Y., Alexeev, D.G., Kostryukova, E.S., Manolov, A.I., Altukhov, I.A., Khazigaleeva, R.A., Seredina, A.V., Kovalchuk, S.I., et al. (2015). Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens*. BMC Plant Biol *15*, 87.

Fialkov, A.B., Steiner, U., Jones, L., Amirav, A. (2007). Sensitivity and noise in GC-MS: Achieveing low limits of detection for difficult analytes. Int. J. Mass Spectrom *260*, 47-53.

Fischer, E., and Sauer, U. (2003). Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. Eur J Biochem *270*, 880-891.

Foreman, J., and Dolan, L. (2001). Root Hairs as a Model System for Studying Plant Cell Growth. Annals of Botany *88*, 1-7.

Fuhrer, T., Fischer, E., and Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. J Bacteriol *187*, 1581-1590.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res *31*, 3784-3788.

Ghosh, A., Nilmeier, J., Weaver, D., Adams, P.D., Keasling, J.D., Mukhopadhyay, A., Petzold, C.J., and Martin, H.G. (2014). A peptide-based method for ^{13}C Metabolic Flux Analysis in microbial communities. PLoS Comput Biol *10*, e1003827.

Gilroy, S., and Jones, D.L. (2000). Through form to function: root hair development and nutrient uptake. Trends Plant Sci *5*, 56-60.

Goto-Inoue, N., Setou, M., and Zaima, N. (2010). Visualization of spatial distribution of gamma-aminobutyric acid in eggplant (Solanum melongena) by matrix-assisted laser desorption/ionization imaging mass spectrometry. Anal Sci *26*, 821-825.

Graham, J.W., Williams, T.C., Morgan, M., Fernie, A.R., Ratcliffe, R.G., and Sweetlove, L.J. (2007). Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. Plant Cell *19*, 3723-3738.

Grassl, J., Taylor, N.L., and Millar, A.H. (2011). Matrix-assisted laser desorption/ionisation mass spectrometry imaging and its development for plant protein imaging. Plant Methods *7*, 21.

Hamm, G., Carre, V., Poutaraud, A., Maunit, B., Frache, G., Merdinoglu, D., and Muller, J.F. (2010). Determination and imaging of metabolites from Vitis vinifera leaves by laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun Mass Spectrom *24*, 335-342.

Haverkorn van Rijsewijk, B.R., Nanchen, A., Nallet, S., Kleijn, R.J., and Sauer, U. (2011). Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli. Mol Syst Biol *7*, 477.

Heise, R., Arrivault, S., Szecowka, M., Tohge, T., Nunes-Nesi, A., Stitt, M., Nikoloski, Z., and Fernie, A.R. (2014). Flux profiling of photosynthetic carbon metabolism in intact plants. Nat Protoc *9*, 1803-1824.

Hemm, M.R., Rider, S.D., Ogas, J., Murry, D.J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J *38*, 765-778.

Horn, P.J., Korte, A.R., Neogi, P.B., Love, E., Fuchs, J., Strupat, K., Borisjuk, L., Shulaev, V., Lee, Y.J., and Chapman, K.D. (2012). Spatial mapping of lipids at cellular resolution in embryos of cotton. Plant Cell *24*, 622-636.

Isermann, N., and Wiechert, W. (2003). Metabolic isotopomer labeling systems. Part II: structural flux identifiability analysis. Math Biosci *183*, 175-214.

Iyer, V.V., Sriram, G., Fulton, D.B., Zhou, R., Westgate, M.E., and Shanks, J.V. (2008). Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons. Plant Cell Environ *31*, 506-517.

Janfelt, C. (2015). Imaging of plant materials using indirect desorption electrospray ionization mass spectrometry. Methods Mol Biol *1203*, 91-97.

Johnson, C.M., Stout, P.R., Broyer, T.C., and Carlton, A.B. (1957). Comparative chlorine requirements of different plant spieces. Plant and Soil *8*, 337-353.

Jorda, J., Jouhten, P., Camara, E., Maaheimo, H., Albiol, J., and Ferrer, P. (2012). Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures. Microb Cell Fact *11*, 57.

Jun, J.H., Song, Z., Liu, Z., Nikolau, B.J., Yeung, E.S., and Lee, Y.J. (2010). High-spatial and high-mass resolution imaging of surface metabolites of Arabidopsis thaliana by laser desorption-ionization mass spectrometry using colloidal silver. Anal Chem *82*, 3255-3265.

Junker, B.H., Lonien, J., Heady, L.E., Rogers, A., and Schwender, J. (2007). Parallel determination of enzyme activities and in vivo fluxes in Brassica napus embryos grown on organic or inorganic nitrogen source. Phytochemistry *68*, 2232-2242.

Kajiyama, T., Fujii, A., Arikawa, K., Habu, T., Mochizuki, N., Nagatani, A., and Kambara, H. (2015). Position-specific gene expression analysis using a microgram dissection method combined with on-bead cDNA library construction. Plant Cell Physiol *56*, 1320-1328.

Karrer, E.E., Lincoln, J.E., Hogenhout, S., Bennett, A.B., Bostock, R.M., Martineau, B., Lucas, W.J., Gilchrist, D.G., and Alexander, D. (1995). In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries. Proc Natl Acad Sci U S A *92*, 3814-3818.

Kaspar, S., Peukert, M., Svatos, A., Matros, A., and Mock, H.P. (2011). MALDI-imaging mass spectrometry - An emerging technique in plant biology. Proteomics *11*, 1840-1850.

Kilburn, M.R., Jones, D.L., Clode, P.L., Cliff, J.B., Stockdale, E.A., Herrmann, A.M., and Murphy, D.V. (2010). Application of nanoscale secondary ion mass spectrometry to plant cell research. Plant Signal Behav *5*, 760-762.

Kruger, N.J., Huddleston, J.E., Le Lay, P., Brown, N.D., and Ratcliffe, R.G. (2007). Network flux analysis: impact of ^{13}C-substrates on metabolism in Arabidopsis thaliana cell suspension cultures. Phytochemistry *68*, 2176-2188.

Kruger, N.J., Masakapalli, S.K., and Ratcliffe, R.G. (2012). Strategies for investigating the plant metabolic network with steady-state metabolic flux analysis: lessons from an Arabidopsis cell culture and other systems. J Exp Bot *63*, 2309-2323.

Kruger, N.J., and Ratcliffe, R.G. (2009). Insights into plant metabolic networks from steady-state metabolic flux analysis. Biochimie *91*, 697-702.

Kruger, N.J., and Ratcliffe, R.G. (2015). Fluxes through plant metabolic networks: measurements, predictions, insights and challenges. Biochem J *465*, 27-38.

Lan, P., Li, W., Lin, W.D., Santi, S., and Schmidt, W. (2013). Mapping gene activity of Arabidopsis root hairs. Genome Biol *14*, R67.

Lee, J.Y., Colinas, J., Wang, J.Y., Mace, D., Ohler, U., and Benfey, P.N. (2006). Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. Proc Natl Acad Sci U S A *103*, 6055-6060.

Lee, M.M., and Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. Cell *99*, 473-483.

Li, H., Shen, J., Zhang, F., Tang, C., and Lambers, H. (2008a). Is there a critical level of shoot phosphorus concentration for cluster-root formation in Lupinus albus? Funct Plant Biol *35*, 328-336.

Li, Y., Shrestha, B., and Vertes, A. (2008b). Atmospheric pressure infrared MALDI imaging mass spectrometry for plant metabolomics. Anal Chem *80*, 407-420.

Libault, M., Brechenmacher, L., Cheng, J., Xu, D., and Stacey, G. (2010). Root hair systems biology. Trends Plant Sci *15*, 641-650.

Lommen, A. (2009). MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem *81*, 3079-3086.

Lonien, J., and Schwender, J. (2009). Analysis of metabolic flux phenotypes for two Arabidopsis mutants with severe impairment in seed storage lipid synthesis. Plant Physiol *151*, 1617-1634.

Ma, Z., Baskin, T.I., Brown, K.M., and Lynch, J.P. (2003). Regulation of root elongation under phosphorus stress involves changes in ethylene responsiveness1. Plant Physiol *131*, 1381-1390.

Mahonen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N., and Helariutta, Y. (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev *14*, 2938-2943.

Mandy, D.E., Goldford, J.E., Yang, H., Allen, D.K., and Libourel, I.G. (2014). Metabolic flux analysis using ^{13}C peptide label measurements. Plant J *77*, 476-486.

Marco-Urrea, E., Seifert, J., von Bergen, M., and Adrian, L. (2012). Stable isotope peptide mass spectrometry to decipher amino acid metabolism in Dehalococcoides strain CBDB1. J Bacteriol *194*, 4169-4177.

Martin, J.C., Maillot, M., Mazerolles, G., Verdu, A., Lyan, B., Migne, C., Defoort, C., Canlet, C., Junot, C., Guillou, C., et al. (2015). Can we trust untargeted metabolomics? Results of the metaboring initiative, a large-scale, multi-instrument inter-laboratory study. Metabolomics *11*, 807-821.

Masakapalli, S.K., Bryant, F.M., Kruger, N.J., and Ratcliffe, R.G. (2014). The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a flexible balance between the cytosolic and plastidic contributions to carbohydrate oxidation in response to phosphate limitation. Plant J *78*, 964-977.

Masakapalli, S.K., Kruger, N.J., and Ratcliffe, R.G. (2013). The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a complex response to changes in nitrogen supply. Plant J *74*, 569-582.

Masakapalli, S.K., Le Lay, P., Huddleston, J.E., Pollock, N.L., Kruger, N.J., and Ratcliffe, R.G. (2010). Subcellular flux analysis of central metabolism in a heterotrophic Arabidopsis cell suspension using steady-state stable isotope labeling. Plant Physiol *152*, 602-619.

Matas, A.J., Yeats, T.H., Buda, G.J., Zheng, Y., Chatterjee, S., Tohge, T., Ponnala, L., Adato, A., Aharoni, A., Stark, R., et al. (2011). Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. Plant Cell *23*, 3893-3910.

Matsuda, F., Wakasa, K., and Miyagawa, H. (2007). Metabolic flux analysis in plants using dynamic labeling technique: application to tryptophan biosynthesis in cultured rice cells. Phytochemistry *68*, 2290-2301.

Moussaieff, A., Rogachev, I., Brodsky, L., Malitsky, S., Toal, T.W., Belcher, H., Yativ, M., Brady, S.M., Benfey, P.N., and Aharoni, A. (2013). High-resolution metabolic mapping of cell types in plant roots. Proc Natl Acad Sci U S A *110*, E1232-1241.

Mullen, A.K., Clench, M.R., Crosland, S., and Sharples, K.R. (2005). Determination of agrochemical compounds in soya plants by imaging matrix-assisted laser desorption/ionisation mass spectrometry. Rapid Commun Mass Spectrom *19*, 2507-2516.

Müller, T., Oradu, S., Ifa, D.R., Cooks, R.G., and Kräutler, B. (2011). Direct plant tissue analysis and imprint imaging by desorption electrospray ionization mass spectrometry. Anal Chem *83*, 5754-5761.

Nakazono, M., Qiu, F., Borsuk, L.A., and Schnable, P.S. (2003). Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell *15*, 583-596.

Nawy, T., Lee, J.Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional profile of the Arabidopsis root quiescent center. Plant Cell *17*, 1908-1925.

Niklas, J., Schneider, K., and Heinzle, E. (2010). Metabolic flux analysis in eukaryotes. Curr Opin Biotechnol *21*, 63-69.

Niu, Y.F., Chai, R.S., Jin, G.L., Wang, H., Tang, C.X., and Zhang, Y.S. (2013). Responses of root architecture development to low phosphorus availability: a review. Ann Bot *112*, 391-408.

Noh, K., and Wiechert, W. (2011). The benefits of being transient: isotope-based metabolic flux analysis at the short time scale. Appl Microbiol Biotechnol *91*, 1247-1265.

Okahashi, N., Kajihata, S., Furusawa, C., and Shimizu, H. (2014). Reliable metabolic flux estimation in Escherichia coli central carbon metabolism using intracellular free amino acids. Metabolites *4*, 408-420.

Paul, M.J., and Pellny, T.K. (2003). Carbon metabolite feedback regulation of leaf photosynthesis and development. J Exp Bot *54*, 539-547.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Petersson, S., Lindén, P., Moritz, T., and Ljung, K. (2015). Cell-type specific metabolic profiling of Arabidopsis thaliana protoplasts as a tool for plant systems biology. Metabolomics, 1-11. doi: 10.1007/s11306-015-0814-7

Petricka, J.J., Schauer, M.A., Megraw, M., Breakfield, N.W., Thompson, J.W., Georgiev, S., Soderblom, E.J., Ohler, U., Moseley, M.A., Grossniklaus, U., et al. (2012a). The protein expression landscape of the Arabidopsis root. Proc Natl Acad Sci U S A *109*, 6811-6818.

Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012c). Control of Arabidopsis root development. Annu Rev Plant Biol *63*, 563-590.

Qiao, Z., and Libault, M. (2013). Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. Front Plant Sci *4*, 484.

Ratcliffe, R.G., and Shachar-Hill, Y. (2005). Revealing metabolic phenotypes in plants: inputs from NMR analysis. Biol Rev Camb Philos Soc *80*, 27-43.

Ratcliffe, R.G., and Shachar-Hill, Y. (2006). Measuring multiple fluxes through plant metabolic networks. Plant J *45*, 490-511.

Rogers, E.D., Jackson, T., Moussaieff, A., Aharoni, A., and Benfey, P.N. (2012). Cell type-specific transcriptional profiling: implications for metabolite profiling. Plant J *70*, 5-17.

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

Roscher, A., Kruger, N.J., and Ratcliffe, R.G. (2000). Strategies for metabolic flux analysis in plants using isotope labelling. J Biotechnol *77*, 81-102.

Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics *7*, 282-289.

Rühl, M., Hardt, W.D., and Sauer, U. (2011). Subpopulation-specific metabolic pathway usage in mixed cultures as revealed by reporter protein-based ^{13}C analysis. Appl Environ Microbiol *77*, 1816-1821.

Rühl, M., Rupp, B., Noh, K., Wiechert, W., Sauer, U., and Zamboni, N. (2012). Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of ^{13}C metabolic flux analysis. Biotechnol Bioeng *109*, 763-771.

Saerens, D., Conrath, K., Govaert, J., and Muyldermans, S. (2008). Disulfide bond introduction for general stabilization of immunoglobulin heavy-chain variable domains. J Mol Biol *377*, 478-488.

Saerens, D., Pellis, M., Loris, R., Pardon, E., Dumoulin, M., Matagne, A., Wyns, L., Muyldermans, S., and Conrath, K. (2005). Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol *352*, 597-607.

Sato, F., Hashimoto, T., Hachiya, A., Tamura, K., Choi, K.B., Morishige, T., Fujimoto, H., and Yamada, Y. (2001). Metabolic engineering of plant alkaloid biosynthesis. Proc Natl Acad Sci U S A *98*, 367-372.

Sauer, U. (2006). Metabolic networks in motion: ^{13}C-based flux analysis. Mol Syst Biol *2*, 62.

Schad, M., Lipton, M.S., Giavalisco, P., Smith, R.D., and Kehr, J. (2005a). Evaluation of two-dimensional electrophoresis and liquid chromatography--tandem mass spectrometry for tissue-specific protein profiling of laser-microdissected plant samples. Electrophoresis *26*, 2729-2738.

Schad, M., Mungur, R., Fiehn, O., and Kehr, J. (2005b). Metabolic profiling of laser microdissected vascular bundles of Arabidopsis thaliana. Plant Methods *1*, 2.

Schiefelbein, J.W. (2000). Constructing a plant cell. The genetic control of root hair development. Plant Physiol *124*, 1525-1531.

Schwarzlander, M., Fricker, M.D., and Sweetlove, L.J. (2009). Monitoring the in vivo redox state of plant mitochondria: effect of respiratory inhibitors, abiotic stress and assessment of recovery from oxidative challenge. Biochim Biophys Acta *1787*, 468-475.

Schwender, J., Goffman, F., Ohlrogge, J.B., and Shachar-Hill, Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature *432*, 779-782.

Schwender, J., and Ohlrogge, J.B. (2002). Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing Brassica napus embryos. Plant Physiol *130*, 347-361.

Schwender, J., Ohlrogge, J.B., and Shachar-Hill, Y. (2003). A flux model of glycolysis and the oxidative pentosephosphate pathway in developing Brassica napus embryos. J Biol Chem *278*, 29442-29453.

Schwender, J., Shachar-Hill, Y., and Ohlrogge, J.B. (2006). Mitochondrial metabolism in developing embryos of Brassica napus. J Biol Chem *281*, 34040-34047.

Schwender, J., Konig, C., Klapperstuck, M., Heinzel, N., Munz, E., Hebbelmann, I., Hay, J.O., Denolf, P., De Bodt, S., Redestig, H., Caestecker, E., Jakob, P.M., Borisjuk, L., Rolletschek, H. (2014) Transcript abundance on its own cannot be used to infer fluxes in central metabolism. Front Plant Sci *5*, 668.

Sena, G., Jung, J.W., and Benfey, P.N. (2004). A broad competence to respond to SHORT ROOT revealed by tissue-specific ectopic expression. Development *131*, 2817-2826.

Shaikh, A.S., Tang, Y.J., Mukhopadhyay, A., and Keasling, J.D. (2008). Isotopomer distributions in amino acids from a highly expressed protein as a proxy for those from total protein. Anal Chem *80*, 886-890.

Shaikh, A.S., Tang, Y.J., Mukhopadhyay, A., Martin, H.G., Gin, J., Benke, P.I., and Keasling, J.D. (2010). Study of stationary phase metabolism via isotopomer analysis of amino acids from an isolated protein. Biotechnol Prog *26*, 52-56.

Shrestha, B., Patt, J.M., and Vertes, A. (2011). In situ cell-by-cell imaging and analysis of small cell populations by mass spectrometry. Anal Chem *83*, 2947-2955.

Shrestha, B., and Vertes, A. (2009). In situ metabolic profiling of single cells by laser ablation electrospray ionization mass spectrometry. Anal Chem *81*, 8265-8271.

Sievert, C., Beuerle, T., Hollmann, J., and Ober, D. (2015). Single cell subtractive transcriptomics for identification of cell-specifically expressed candidate genes of pyrrolizidine alkaloid biosynthesis. Phytochemistry *117*, 17-24.

Simpson, R.J. (2010). Rapid coomassie blue staining of protein gels. Cold Spring Harb Protoc *2010*, 5413.

Smith, R.A., Schuetz, M., Roach, M., Mansfield, S.D., Ellis, B., and Samuels, L. (2013). Neighboring parenchyma cells contribute to Arabidopsis xylem lignification, while lignification of interfascicular fibers is cell autonomous. Plant Cell *25*, 3988-3999.

Sparbier, K., Lange, C., Jung, J., Wieser, A., Schubert, S., and Kostrzewa, M. (2013). MALDI biotyper-based rapid resistance detection by stable-isotope labeling. J Clin Microbiol *51*, 3741-3748.

Spielbauer, G., Margl, L., Hannah, L.C., Romisch, W., Ettenhuber, C., Bacher, A., Gierl, A., Eisenreich, W., and Genschel, U. (2006). Robustness of central carbohydrate metabolism in developing maize kernels. Phytochemistry *67*, 1460-1475.

Sriram, G., Fulton, D.B., Iyer, V.V., Peterson, J.M., Zhou, R., Westgate, M.E., Spalding, M.H., and Shanks, J.V. (2004). Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing. Plant Physiol *136*, 3043-3057.

Sriram, G., Fulton, D.B., and Shanks, J.V. (2007). Flux quantification in central carbon metabolism of *Catharanthus roseus* hairy roots by ^{13}C labeling and comprehensive bondomer balancing. Phytochemistry *68*, 2243-2257.

Sweetlove, L.J., Fell, D., and Fernie, A.R. (2008). Getting to grips with the plant metabolic network. Biochem J *409*, 27-41.

Sweetlove, L.J., Last, R.L., and Fernie, A.R. (2003). Predictive metabolic engineering: a goal for systems biology. Plant Physiol *132*, 420-425.

Szyperski, T. (1995). Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. Eur J Biochem *232*, 433-448.

Takahashi, K., Kozuka, T., Anegawa, A., Nagatani, A., and Mimura, T. (2015). Development and application of a high-resolution imaging mass spectrometer for the study of plant tissues. Plant Cell Physiol *56*, 1329-1338.

Tanaka, N., Kato, M., Tomioka, R., Kurata, R., Fukao, Y., Aoyama, T., and Maeshima, M. (2014a). Characteristics of a root hair-less line of Arabidopsis thaliana under physiological stresses. J Exp Bot *65*, 1497-1512.

Tanaka, S., Fujita, Y., Parry, H.E., Yoshizawa, A.C., Morimoto, K., Murase, M., Yamada, Y., Yao, J., Utsunomiya, S.I., Kajihara, S., et al. (2014b). Mass++: A visualization and analysis tool for mass spectrometry. J Proteome Res. doi: 10.1021/pr500155z.

Tata, A., Perez, C.J., Hamid, T.S., Bayfield, M.A., and Ifa, D.R. (2015). Analysis of metabolic changes in plant pathosystems by imprint imaging DESI-MS. J Am Soc Mass Spectrom *26*, 641-648.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J *37*, 914-939.

To, J.P., Reiter, W.D., and Gibson, S.I. (2002). Mobilization of seed storage lipid by Arabidopsis seedlings is retarded in the presence of exogenous sugars. BMC Plant Biol *2*, 4.

Troufflard, S., Roscher, A., Thomasset, B., Barbotin, J.N., Rawsthorne, S., and Portais, J.C. (2007). In vivo ^{13}C NMR determines metabolic fluxes and steady state in linseed embryos. Phytochemistry *68*, 2341-2350.

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Piques, M.C., Steinhauser, D., et al. (2005). Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. Plant Physiol *138*, 1195-1204.

Usadel, B., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ *32*, 1211-1229.

van Winden, W.A., Heijnen, J.J., Verheijen, P.J., and Grievink, J. (2001). A priori analysis of metabolic flux identifiability from ^{13}C-labeling data. Biotechnol Bioeng *74*, 505-516.

Velickovic, D., Ropartz, D., Guillon, F., Saulnier, L., and Rogniaux, H. (2014). New insights into the structural and spatial variability of cell-wall polysaccharides during wheat grain development, as revealed through MALDI mass spectrometry imaging. J Exp Bot *65*, 2079-2091.

Vrkoslav, V., Muck, A., Cvacka, J., and Svatos, A. (2010). MALDI imaging of neutral cuticular lipids in insects and plants. J Am Soc Mass Spectrom *21*, 220-231.

Wahl, S.A., Dauner, M., and Wiechert, W. (2004). New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng *85*, 259-268.

Wasylenko, T.M., and Stephanopoulos, G. (2013). Kinetic isotope effects significantly influence intracellular metabolite ^{13}C labeling patterns and flux determination. Biotechnol J *8*, 1080-1089.

Wiechert, W. (2001). ^{13}C metabolic flux analysis. Metab Eng *3*, 195-206.

Wiechert, W., and de Graaf, A.A. (1997). Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng *55*, 101-117.

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

Wiechert, W., Mollney, M., Petersen, S., and de Graaf, A.A. (2001). A universal framework for ^{13}C metabolic flux analysis. Metab Eng *3*, 265-283.

Williams, T.C., Miguet, L., Masakapalli, S.K., Kruger, N.J., Sweetlove, L.J., and Ratcliffe, R.G. (2008). Metabolic network fluxes in heterotrophic Arabidopsis cells: stability of the flux distribution under different oxygenation conditions. Plant Physiol *148*, 704-718.

Williams, T.C., Poolman, M.G., Howden, A.J., Schwarzlander, M., Fell, D.A., Ratcliffe, R.G., and Sweetlove, L.J. (2010). A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. Plant Physiol *154*, 311-323.

Wittmann, C. (2007). Fluxome analysis using GC-MS. Microb Cell Fact *6*, 6.

Xiong, W., Liu, L., Wu, C., Yang, C., and Wu, Q. (2010). ^{13}C-tracer and gas chromatography-mass spectrometry analyses reveal metabolic flux distribution in the oleaginous microalga *Chlorella protothecoides*. Plant Physiol *154*, 1001-1011.

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

Yewdell, J.W., Lacsina, J.R., Rechsteiner, M.C., and Nicchitta, C.V. (2011). Out with the old, in with the new? Comparing methods for measuring protein degradation. Cell Biol Int *35*, 457-462.

Yi, K., Menand, B., Bell, E., and Dolan, L. (2010). A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat Genet *42*, 264-267.

You, L., Liu, H., Blankenship, R.E., and Tang, Y.J. (2015). Using photosystem I as a reporter protein for ^{13}C analysis in a coculture containing cyanobacterium and a heterotrophic bacterium. Anal Biochem *477*, 86-88.

Young, J.D., Shastri, A.A., Stephanopoulos, G., and Morgan, J.A. (2011). Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis. Metab Eng *13*, 656-665.

Zamboni, N., Fendt, S.M., Rühl, M., and Sauer, U. (2009). ^{13}C-based metabolic flux analysis. Nat Protoc *4*, 878-892.

Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M., Burnett, W.J., Logan, K.O., Wang, D., Wang, X., et al. (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell *27*, 513-531.

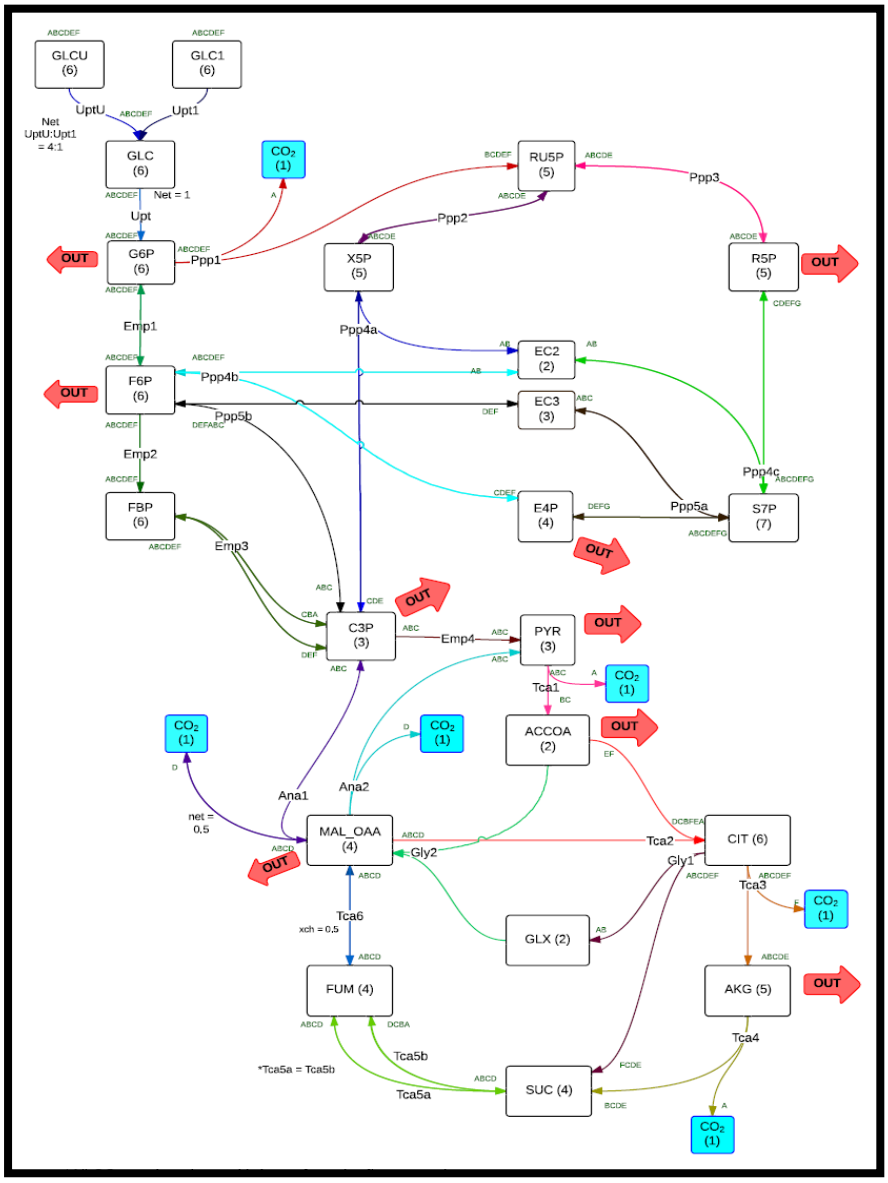
Appendices:

Appendix 1: **Variation in the flux estimates for the second free flux in model containing one free net flux and one free exchange flux, when the flux held constant varies from the set value.**

Free flux held constant	*Free flux varied*	*Dataset used for deducing flux*	*Flux value*	*Estimated mean for flux held constant*
R1_xch	R1	Individual dataset	0	0.358
R1_xch	R1	Individual dataset	0.5	0.488
R1_xch	R1	Individual dataset	0.5	0.496
R1_xch	R1	Average dataset	0.5	0.489
R1_xch	R1	Individual dataset	0.5	0.401
R1_xch	R1	Average dataset	0.5	0.004
R1_xch	R1	Individual dataset	0.99	0.971
R1_xch	R1	Average dataset	0.99	0.962
R1_xch	R1	Individual dataset	0.99	0.401
R1_xch	R1	Average dataset	0.99	0.741
R1	R1_xch	Individual dataset	1	0.020
R1	R1_xch	Individual dataset	1	0.44
R1	R1_xch	Individual dataset	0.5	0.499
R1	R1_xch	Average dataset	0.5	0.523
R1	R1_xch	Average dataset	0.5	0.254
R1	R1_xch	Average dataset	0	0.121

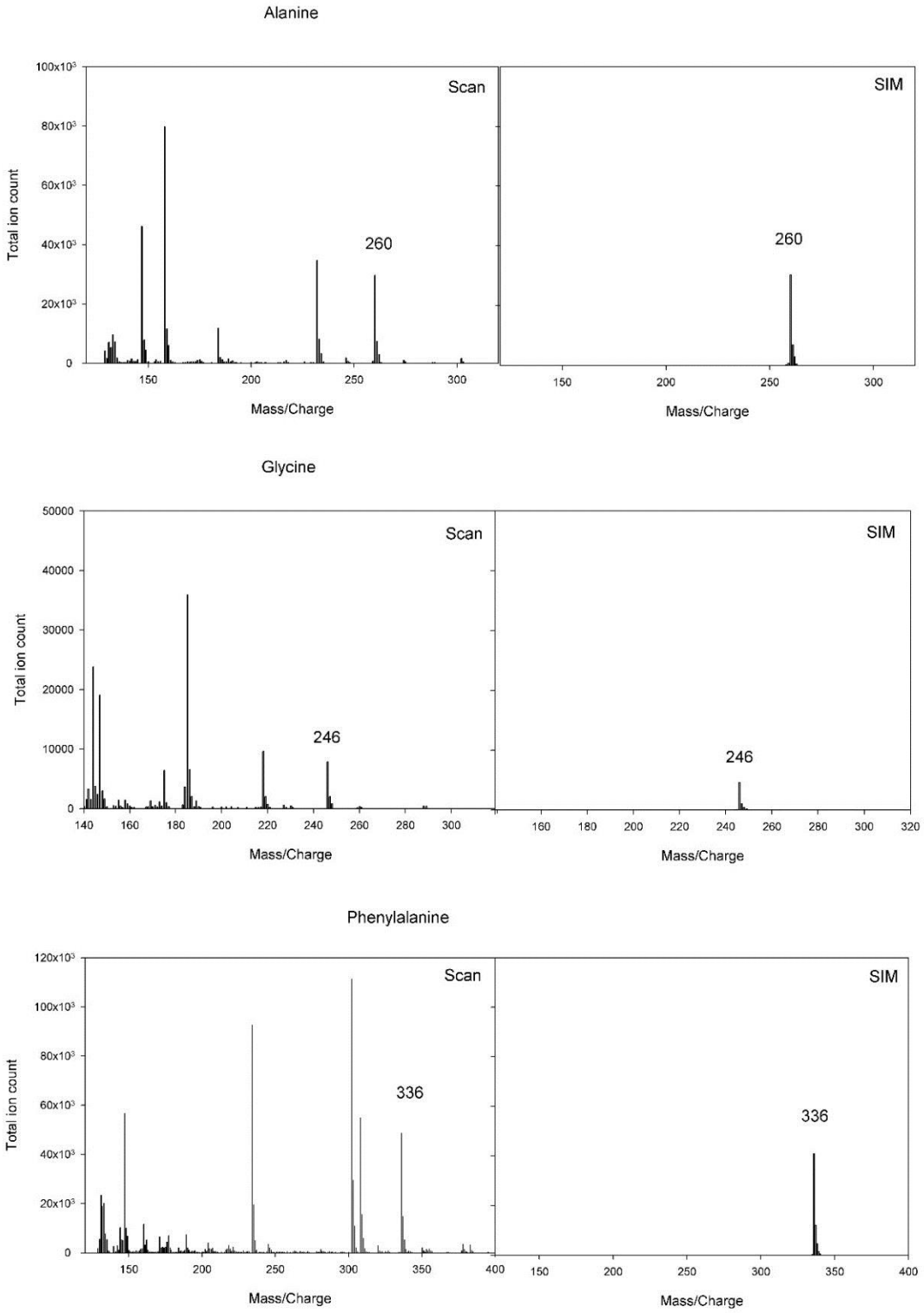
R1	R1_xch	Average dataset	0	0.254

Appendix 2: **Global best fit solutions obtained for a larger more realistic metabolic network containing equal contribuitons from two metabolically distinct cell types show differences to the aggregate fluxes.** Deduced fluxes are the mean ± SD of 100 simulations beginning from random positions within the permissible flux-space. The deduced fluxes differed from the expected fluxesirrespective of whether the contributing networks differed in either internal or output fluxes (Chaiyakorn Srisakvarakul and Nicholas Kruger, unpublished data).

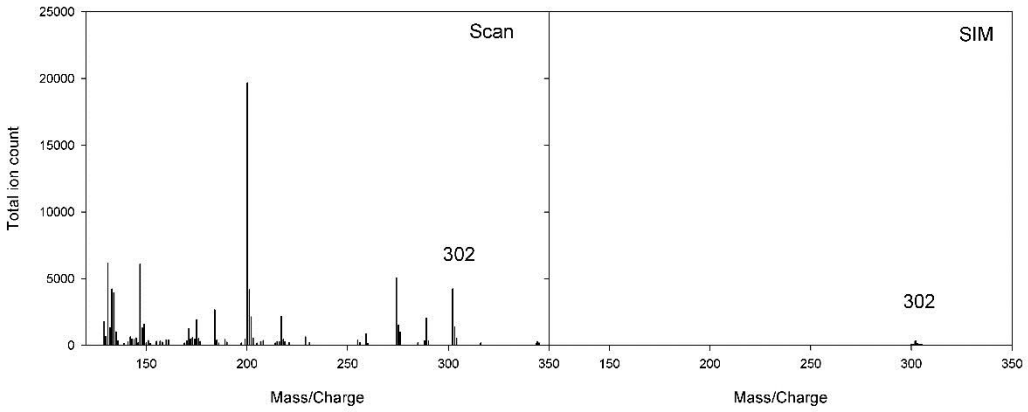


Network variations	Reaction step	Actual flux	Deduced flux
Internal fluxes only	$Ppp1_{net}$	0.500 (0.197; 0.803)	0.487 ± 0.012
	$Gly1_{net}$	0.206 (0.015; 0.397)	0.245 ± 0.015
External and internal fluxes	$Ppp1_{net}$	0.181 (0.121; 0.242)	0.204 ± 0.007
	$Gly1_{net}$	0.059 (0.015; 0.102)	-0.051 ± 0.006

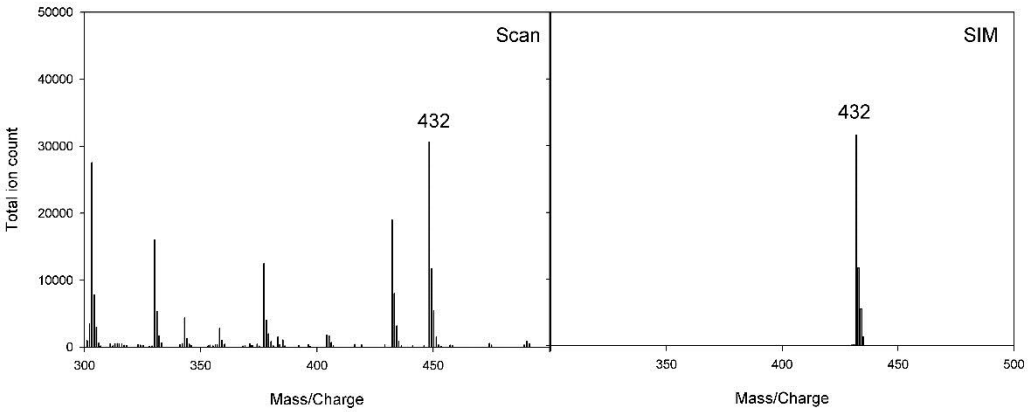
Appendix 3: **Mass spectra of the amino acid fragments detected from GFP immunopurified from a single culture of 30-50 seedlings and prepared using the conventional conditions and protocols.** Each pair of panels presents profiles from analyses performed in scan and SIM mode on the same sample.



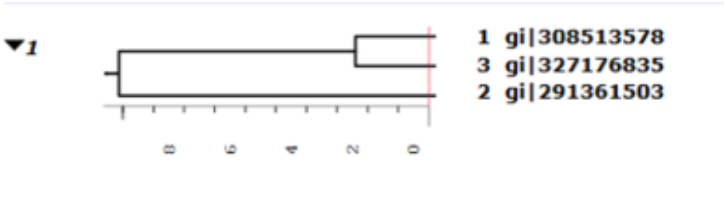
Isoleucine



Glutamate



Appendix 4**: Results for analysis by peptide mass fingerprinting to confirm identity of protein purified with GFP nanotrap**. A trypsin digested sample of purified plant GFP run on a SDS-PAGE gel was sent for analysis by peptide mass fingerprinting at the University of Oxford Central Proteomics Facility. The MASCOT database search results indicate significant matches ($p<0.05$ (Perkins et al., 1999)) with GFP variants with several peptide fragments matching in each case.

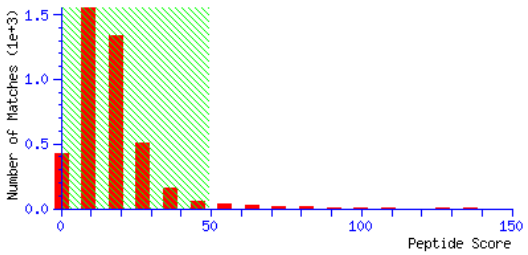


451 GFP [Cloning vector pBAV1KT5gfp]
202 green fluorescent protein GM2 variant [Cloning vector pGBT-PX] >gi|327176848|gb|AEA29859.1|
411 green fluorescent protein [Cloning vector pEF-F3-EGFP-F3-mcs] >gi|336396635|gb|AEI54555.1|

▼*Search parameters*

Type of search	: MS/MS Ion Search
Enzyme	: Trypsin
Variable modifications	: Acetyl (Protein N-term), Deamidated (
Mass values	: Monoisotopic
Protein mass	: Unrestricted
Peptide mass tolerance	: ± 20 ppm
Fragment mass tolerance	: ± 0.5 Da
Max missed cleavages	: 3
Instrument type	: Default
Number of queries	: 4,733

▼*Score distribution*



Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event.
There are **101** peptide matches above identity threshold and **145** matches above homology threshold for **4,733** queries. Histogram score range is (0, 136). On average, individual ions scores **> 49** (beyond green shading) indicate **identity or extensive homology** ($p<0.05$).

Select Summary Report

Format As | Select Summary (protein hits) | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ○ MudPIT scoring ◉ | Ions score or expect cut-off 20 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Require bold red ☐

Preferred taxonomy All entries

Re-Search ◉ All queries ○ Unassigned ○ Below homology threshold ○ Below identity threshold

1. gi|308513578 Mass: 26880 Score: 451 Matches: 22(15) Sequences: 6(6) emPAI: 1.71
GFP [Cloning vector pBAV1KT5gfp]

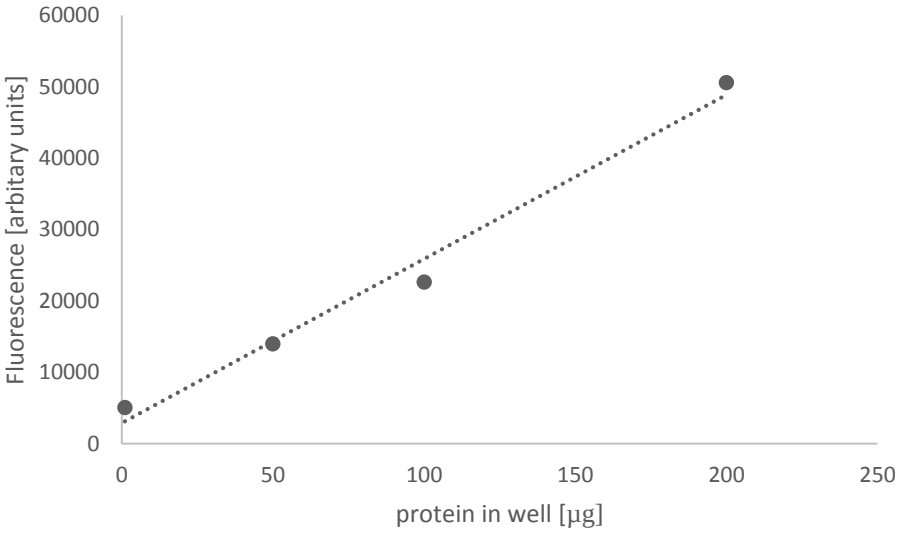
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1322	322.1663	963.4771	963.4774	-0.25	0	(54)	0.011	1		K.EDGNILGHK.L
1323	482.7462	963.4778	963.4774	0.48	0	(41)	0.21	1		K.EDGNILGHK.L
1416	328.1697	981.4874	981.4879	-0.58	0	(41)	0.26	1		K.EDGNILGHK.L
1417	491.7516	981.4886	981.4879	0.64	0	56	0.0074	1		K.EDGNILGHK.L
1777	525.7643	1049.5141	1049.5142	-0.04	0	56	0.0086	1		K.FEGDTLVNR.I 1778
2644	633.7922	1265.5699	1265.5710	-0.91	0	56	0.01	1		K.SAMPEGYVQER.T
2692	641.7896	1281.5647	1281.5659	-0.97	0	(37)	0.88	1		K.SAMPEGYVQER.T
3203	752.3337	1502.6529	1502.6525	0.26	0	79	7.6e-05	1		K.FSVSGEGEGDATYGK.L 3202 3204 3205 3206
4207	1219.1321	2436.2497	2436.2537	-1.62	0	78	0.00017	2		K.GEELFTGVVPILVELDGDVNGHK.F
4208	813.0920	2436.2543	2436.2537	0.26	0	(56)	0.024	2		K.GEELFTGVVPILVELDGDVNGHK.F 4209 4210
4212	1219.6265	2437.2384	2437.2377	0.30	0	(66)	0.0024	1		K.GEELFTGVVPILVELDGDVNGHK.F
4213	813.4209	2437.2410	2437.2377	1.34	0	(56)	0.028	1		K.GEELFTGVVPILVELDGDVNGHK.F
4293	855.7903	2564.3492	2564.3486	0.21	1	63	0.0033	1	U	R.KGEELFTGVVPILVELDGDVNGHK.F
4294	642.0954	2564.3524	2564.3486	1.48	1	(31)	5	1	U	R.KGEELFTGVVPILVELDGDVNGHK.F
4295	856.1173	2565.3301	2565.3326	-1.00	1	(60)	0.0099	1	U	R.KGEELFTGVVPILVELDGDVNGHK.F

Protein sequence coverage: 31%

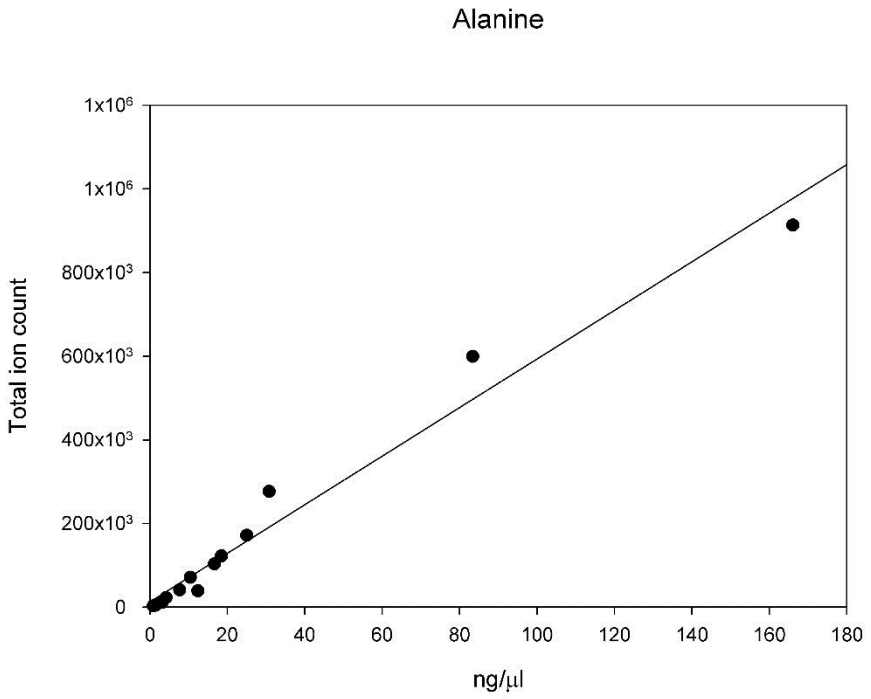
Matched peptides shown in ***bold red***.

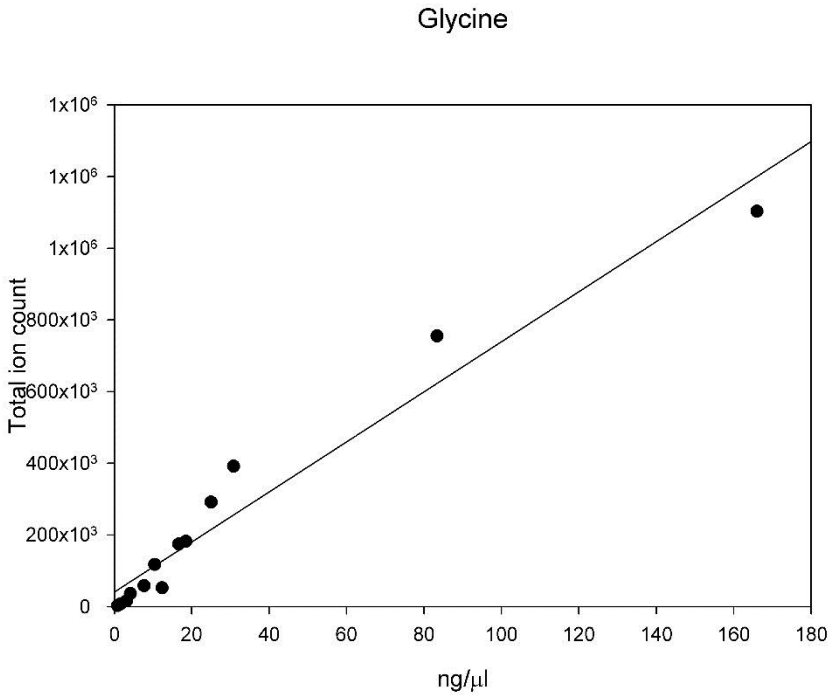
```
  1 MRKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT
 51 GKLPVPWPTL VTTLTYGVQC FSRYPDHMKR YDFFKSAMPE GYVQERTISF
101 KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV
151 YITADKQKNG IKANFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY
201 LSTQSALLKD PNEKRDHMVL LEFVTAAGIT HGMDELYK
```

Appendix 5: **Fluorescence measurements using GFP standard expressed in *Escherichia coli*.** GFP standard of a known concentration (confirmed with Bradford assay using BSA as a standard) was used to measure fluorescence readings (in arbitrary units) from a range of concentrations of GFP. Measurements were taken from a 96-well microplate suitable for measuring fluorescence and sample volume was adjusted to the same total volume. Shown here are measurements from 1 to 200 micrograms. Each measurement was taken three times. Standard deviations are covered by the symbols.

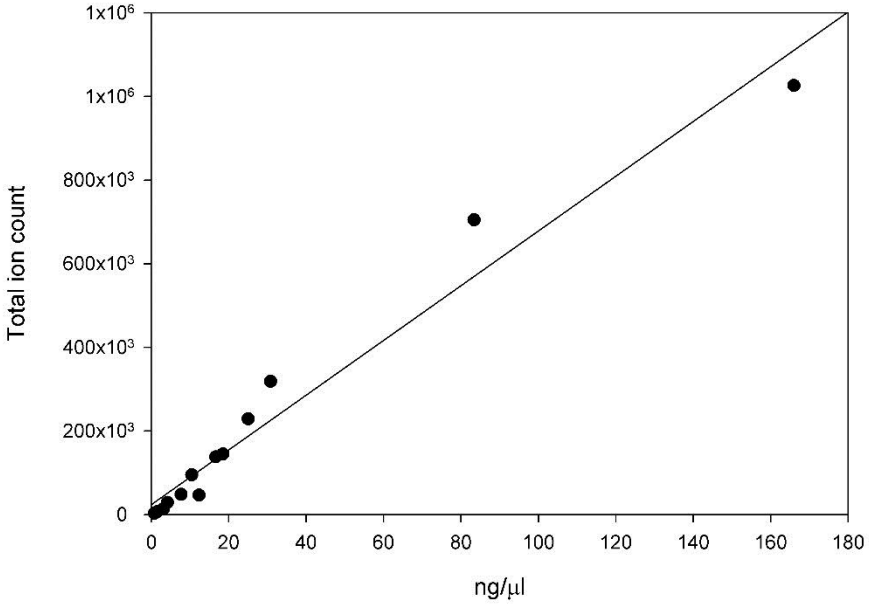


Appendix 6: **Chromatographic analysis of amino acid standards.** Analysis of amino acid standards at different concentrations and the corresponding total ion counts. The trendline represents a linearity, where the relationship between total ion count and concentration is linear.

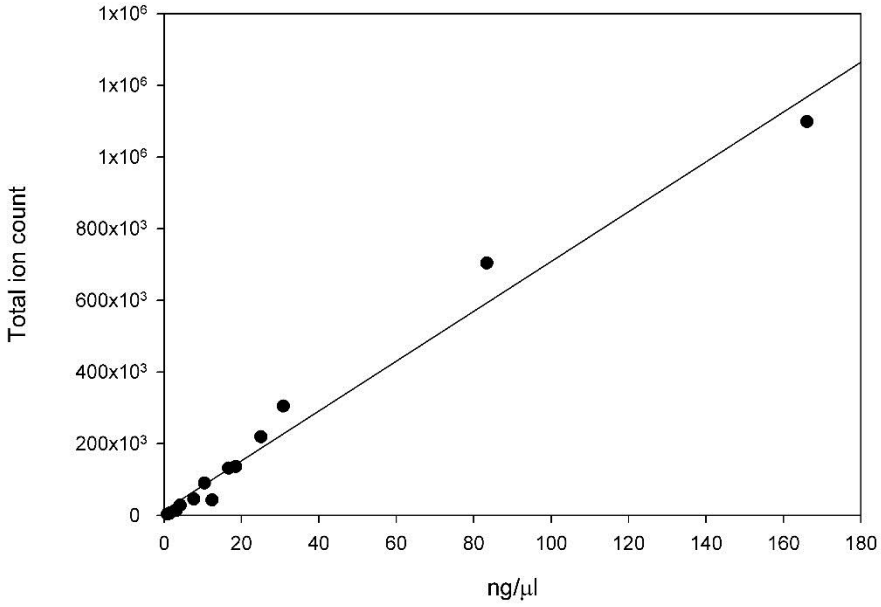




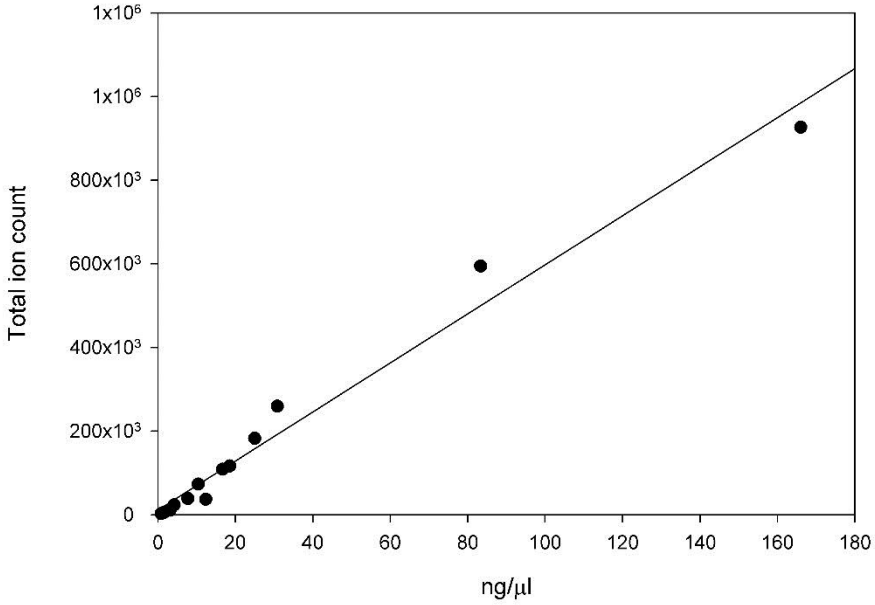
Valine



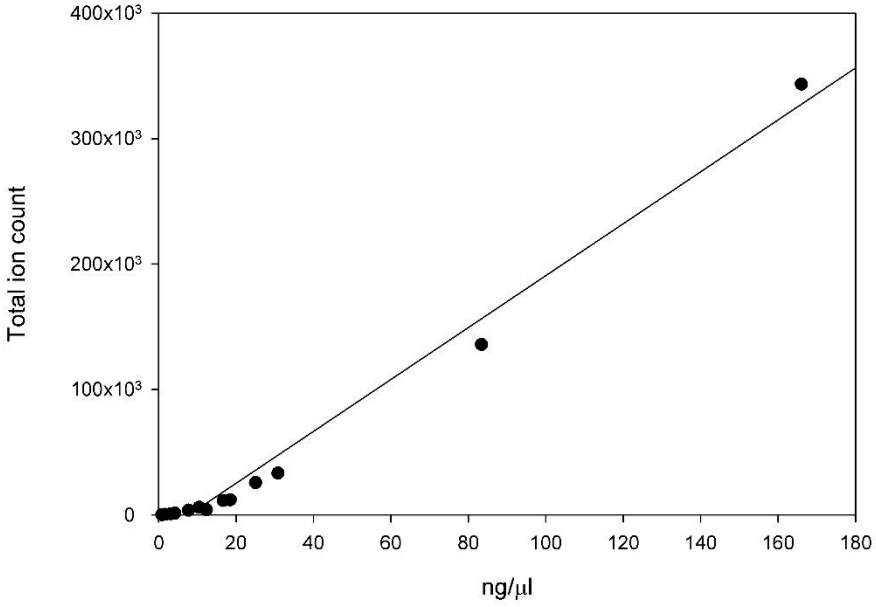
Leucine



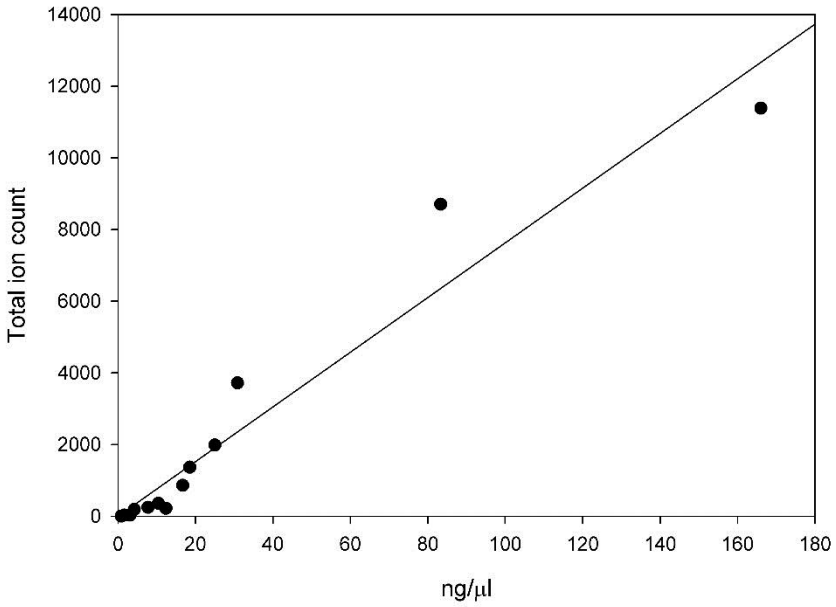
Isoleucine



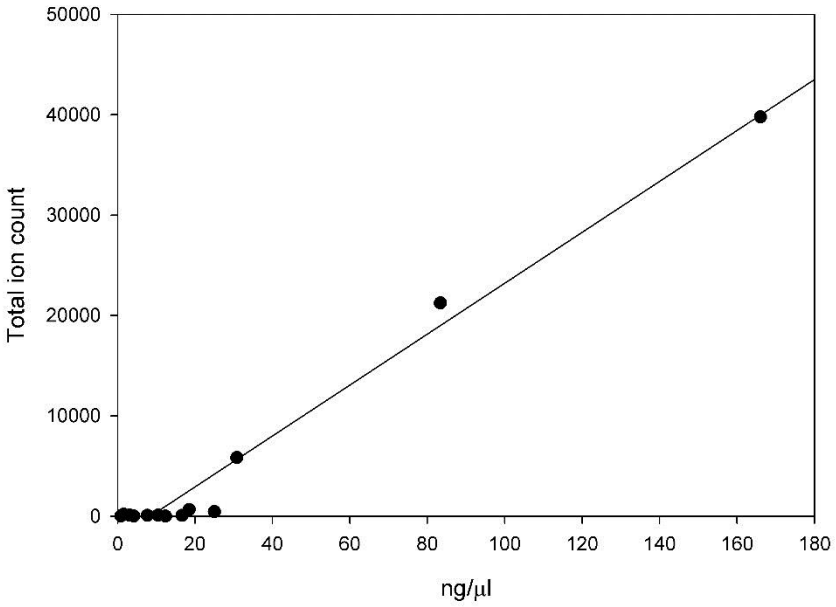
Proline



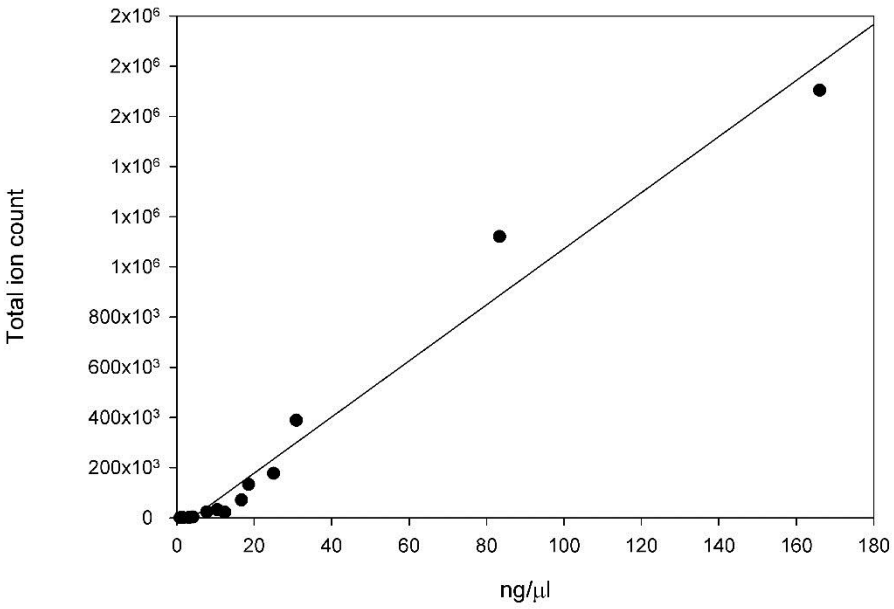
Arginine



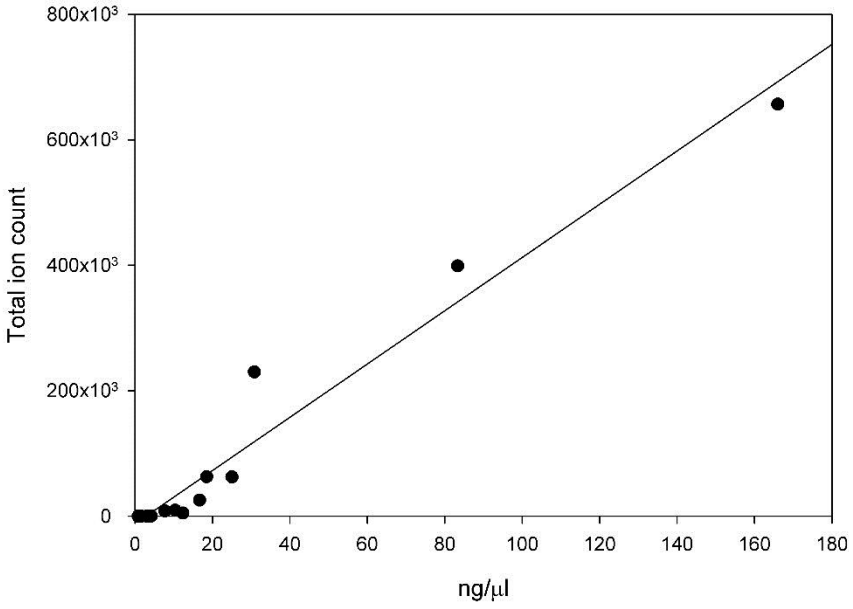
Histidine



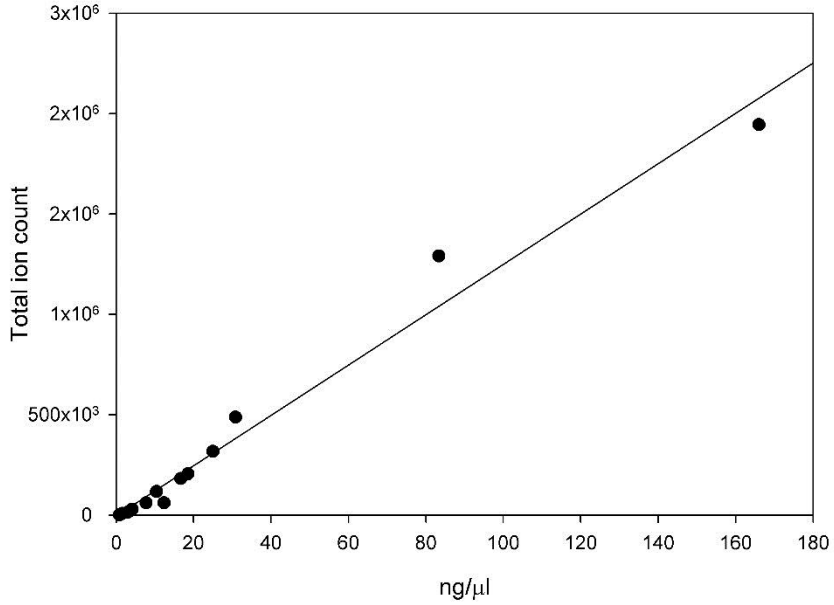
Asparagine



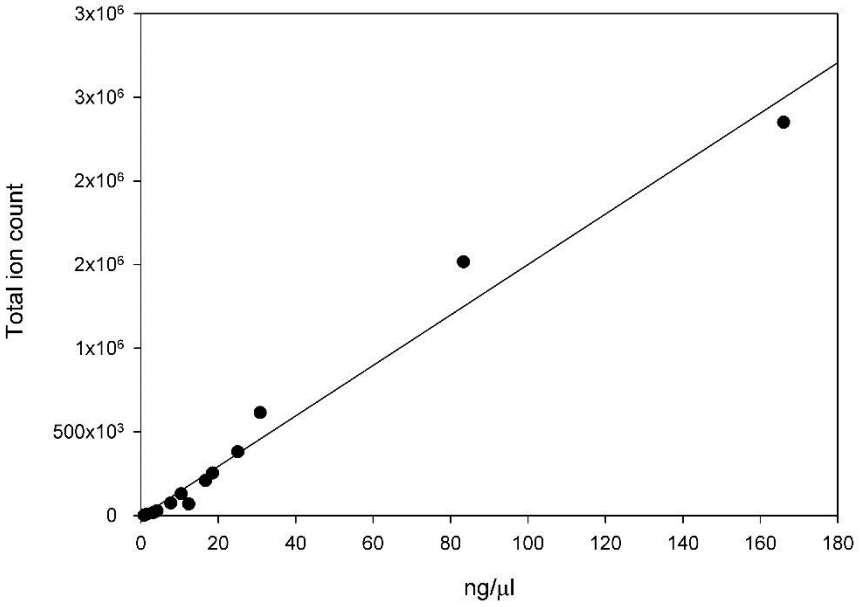
Glutamine



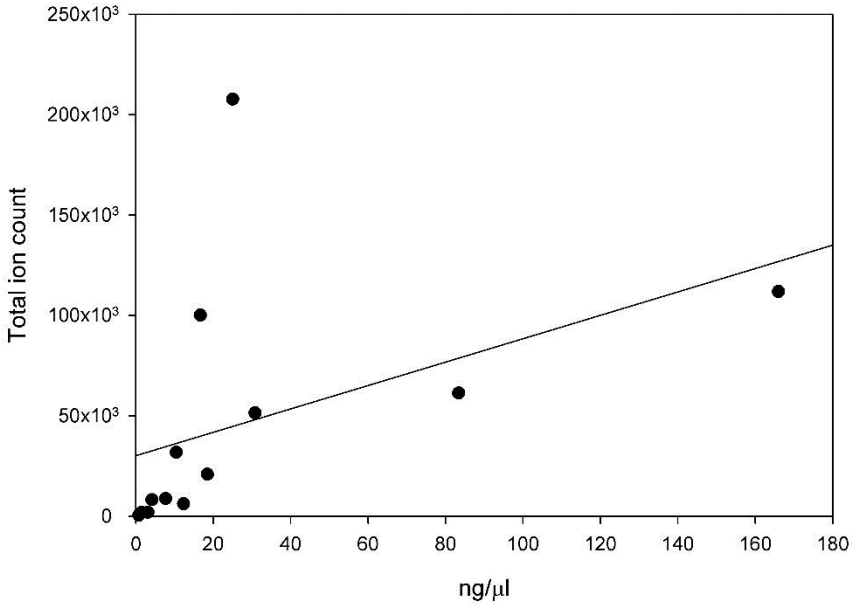
Aspartate



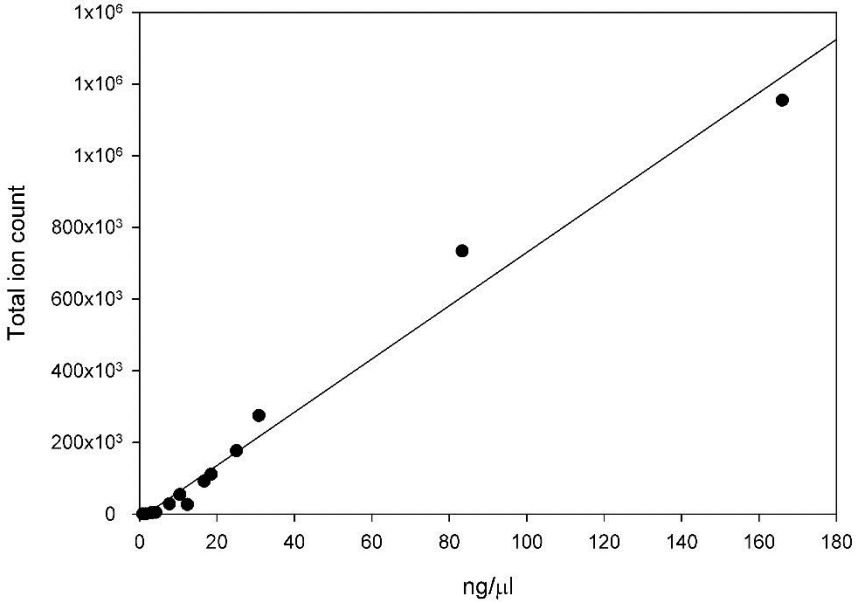
Glutamate



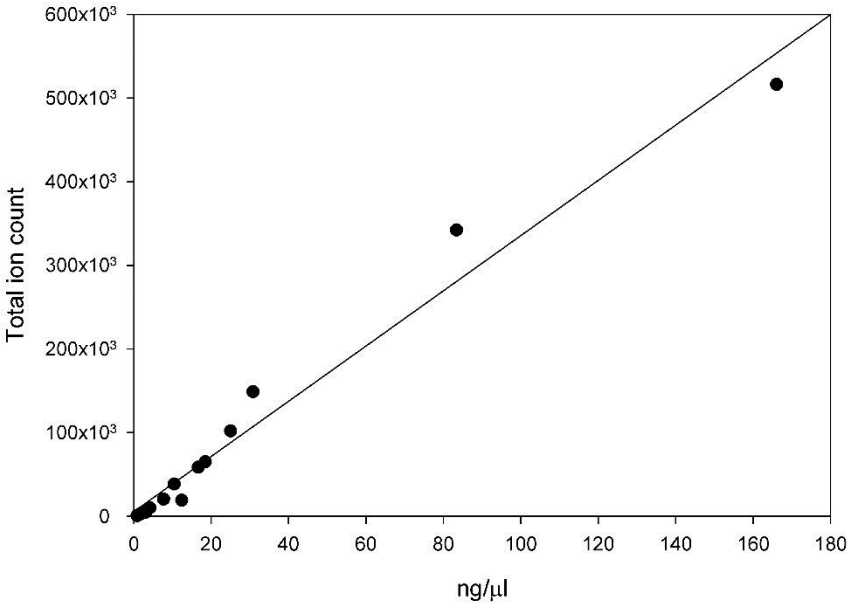
Cysteine



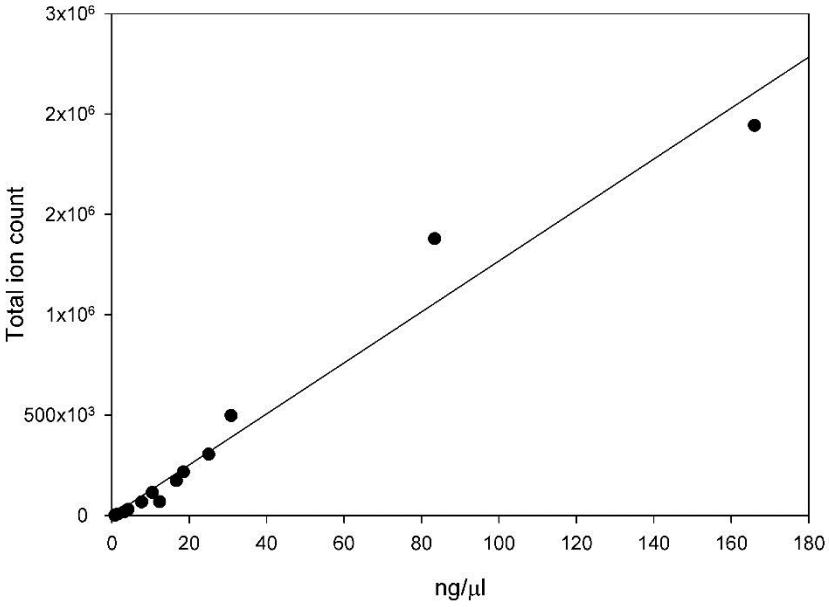
Tyrosine



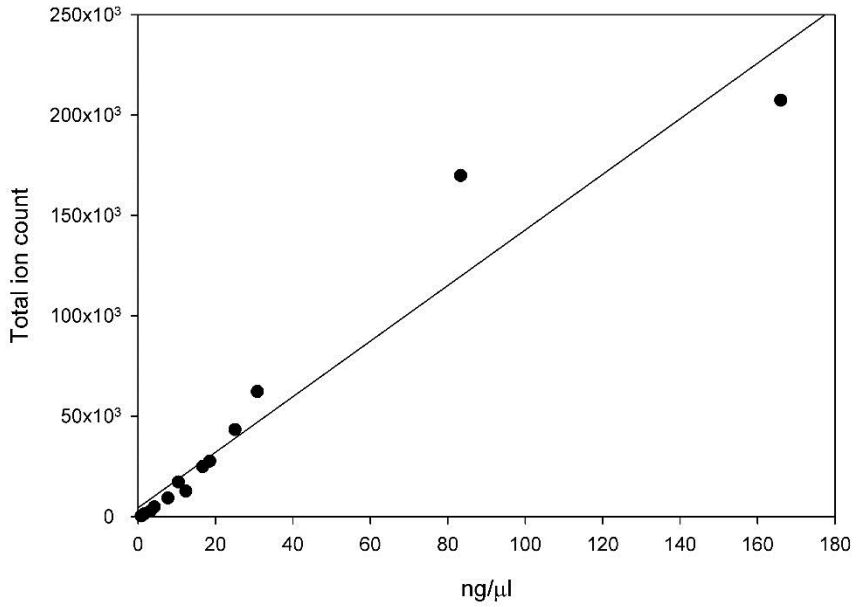
Methionine



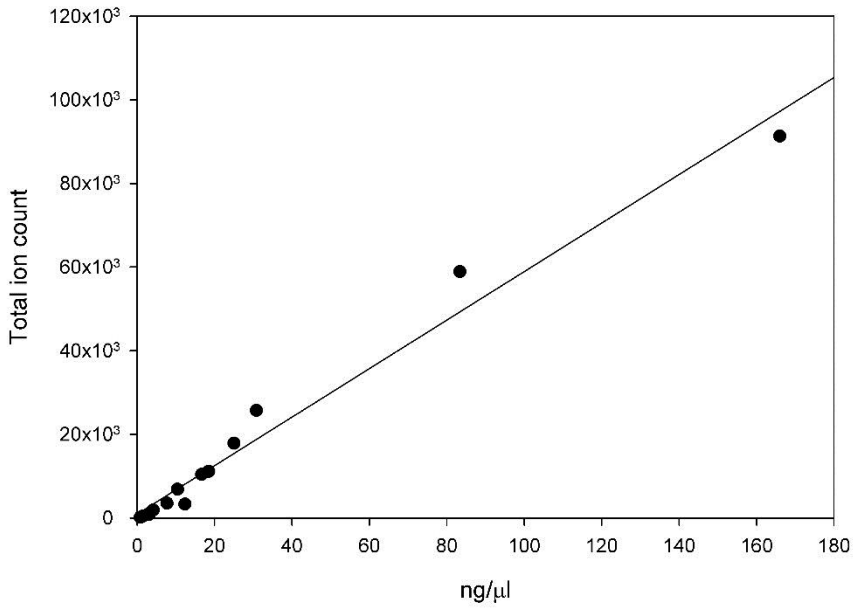
Serine

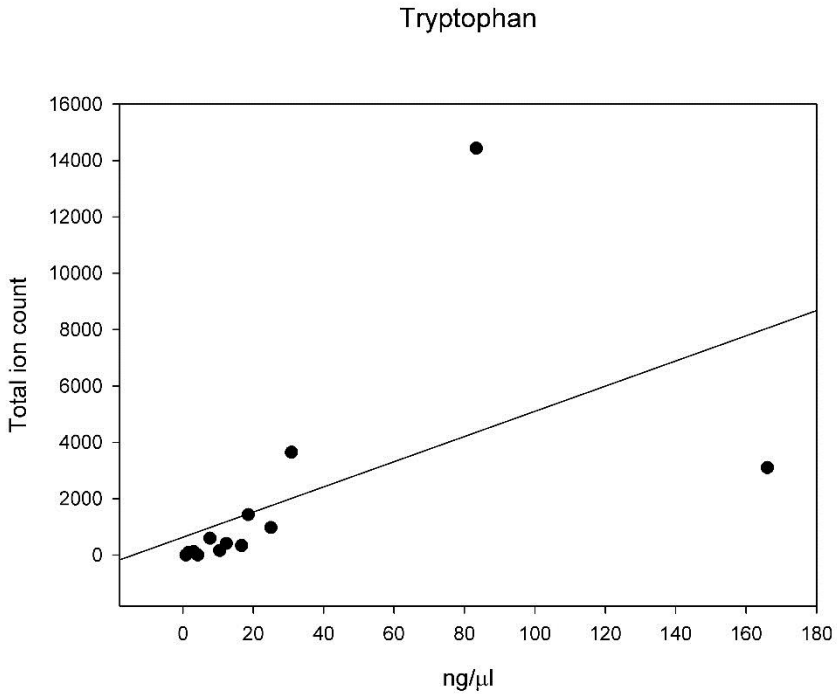


Threonine



Phenylalanine




Tryptophan
Total ion count
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
140
160
180
ng/μl

Appendix 7: **Amino acid composition of GFP**

```
  10         20         30         40         50
MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT
        60         70         80         90        100
GKLPVPWPTL VTTFSYGVQC FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF
       110        120        130        140        150
KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV
       160        170        180        190        200
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY
       210        220        230
LSTQSALSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK
```

Number of amino acids: 238
Molecular weight: 26886.3
Theoretical pI: 5.67

Ala (A)	8	3.4%
Arg (R)	6	2.5%
Asn (N)	13	5.5%
Asp (D)	18	7.6%
Cys (C)	2	0.8%
Gln (Q)	8	3.4%
Glu (E)	16	6.7%
Gly (G)	22	9.2%
His (H)	10	4.2%
Ile (I)	12	5.0%
Leu (L)	19	8.0%
Lys (K)	20	8.4%
Met (M)	6	2.5%
Phe (F)	13	5.5%
Pro (P)	10	4.2%
Ser (S)	11	4.6%
Thr (T)	15	6.3%
Trp (W)	1	0.4%
Tyr (Y)	11	4.6%
Val (V)	17	7.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 34
Total number of positively charged residues (Arg + Lys): 26

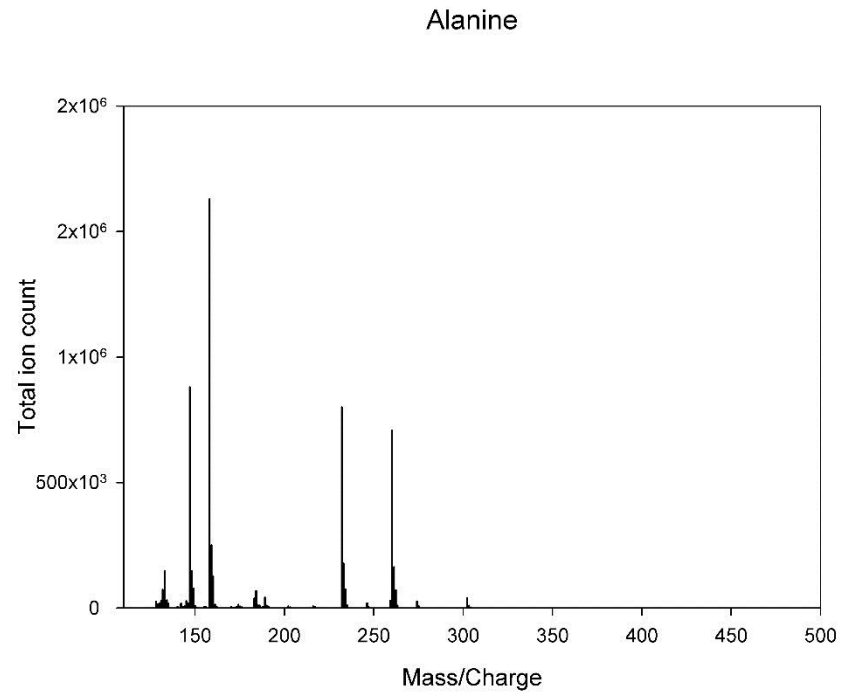
Atomic composition:

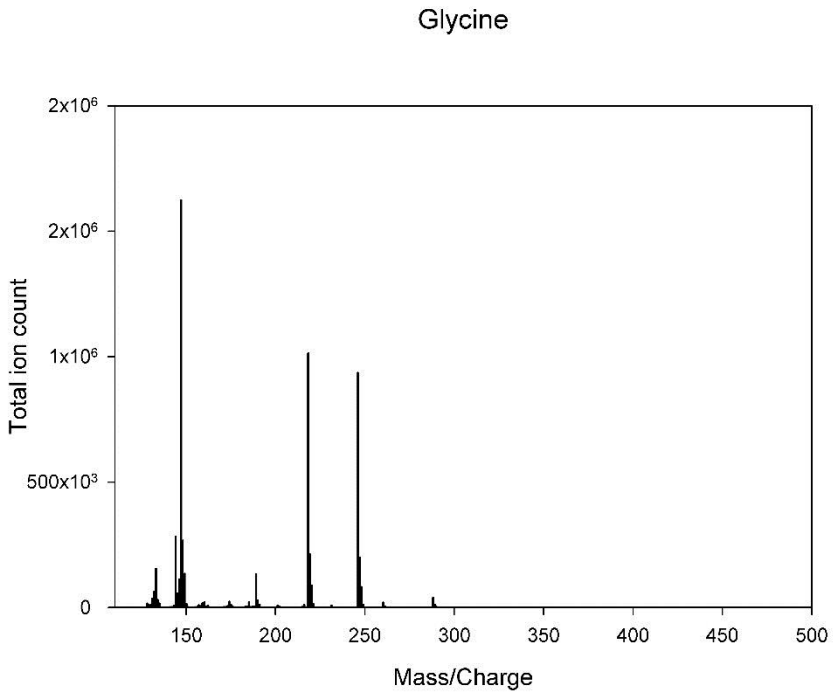
Carbon	C	1205
Hydrogen	H	1848
Nitrogen	N	318
Oxygen	O	365
Sulfur	S	8

Formula: $C_{1205}H_{1848}N_{318}O_{365}S_8$
Total number of atoms: 3744

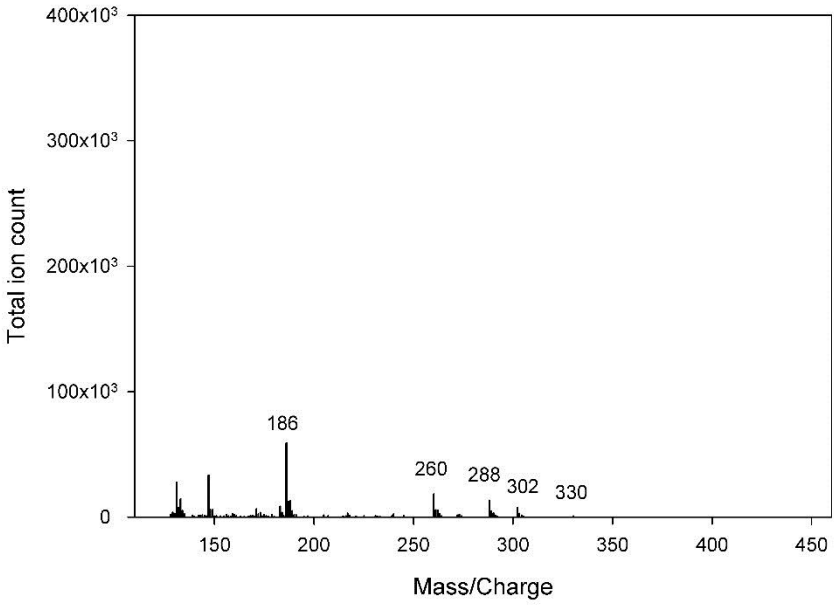
Obtained from the ExPaSy server (Gasteiger et al., 2003).

Appendix 8: **Mass spectra of amino acids derived from a GFP sample immunopurified from three cultures of 30-50 seedlings with constitutive expression and prepared with conventional volumes and conditions.**

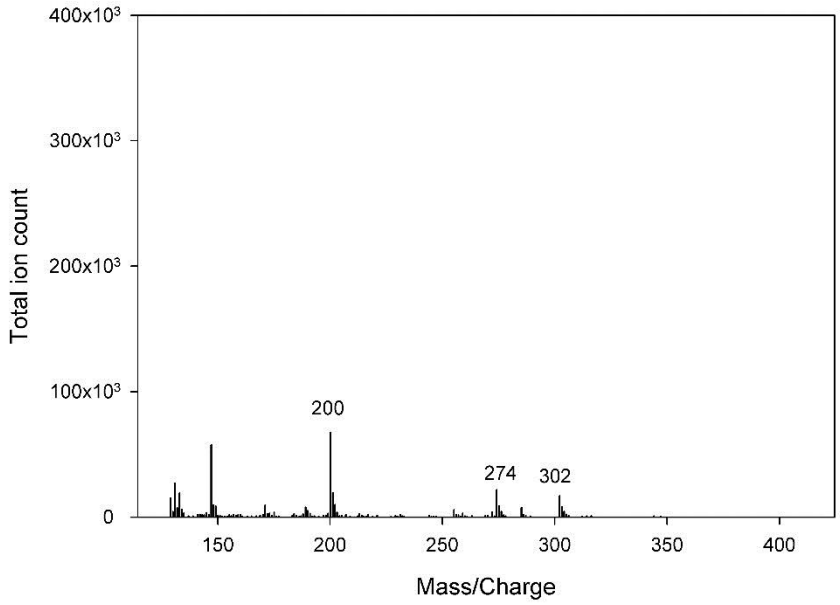




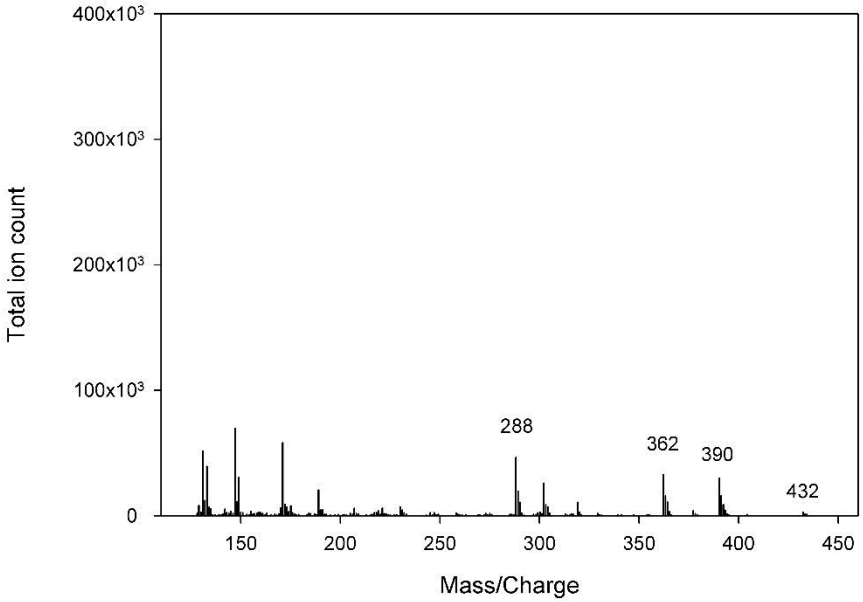
Valine



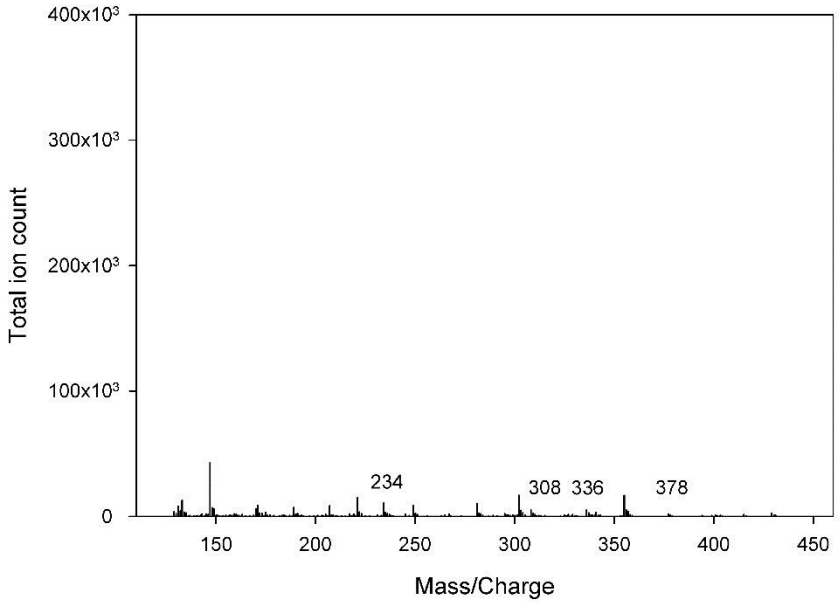
Leucine or isoleucine



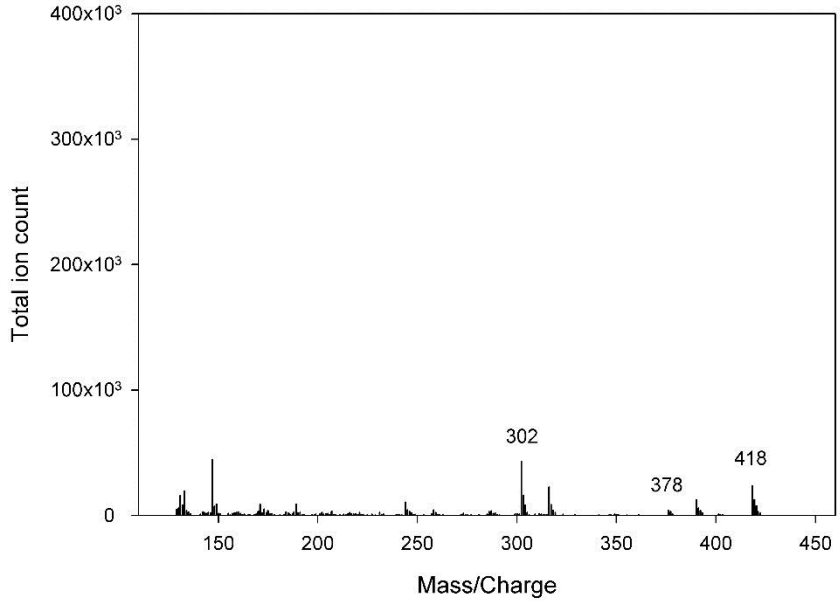
Serine



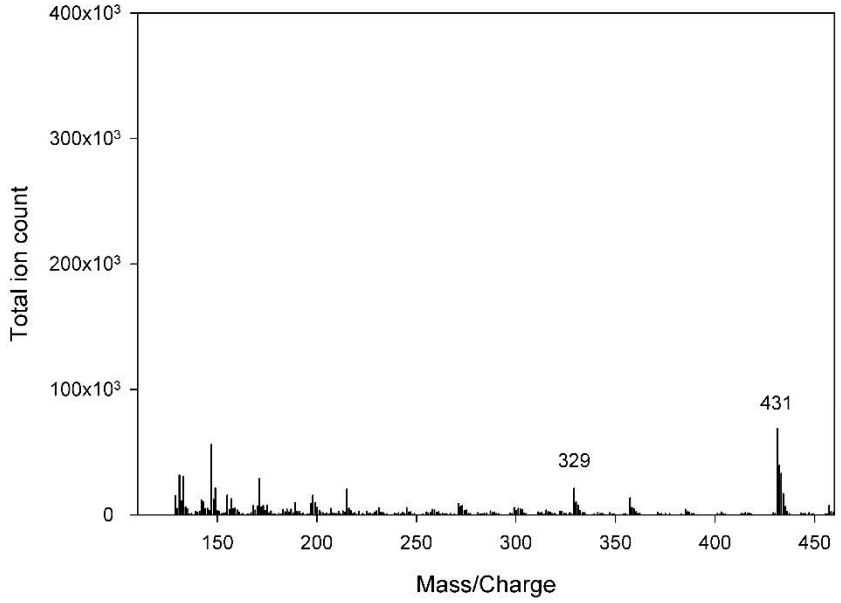
Phenylalanine



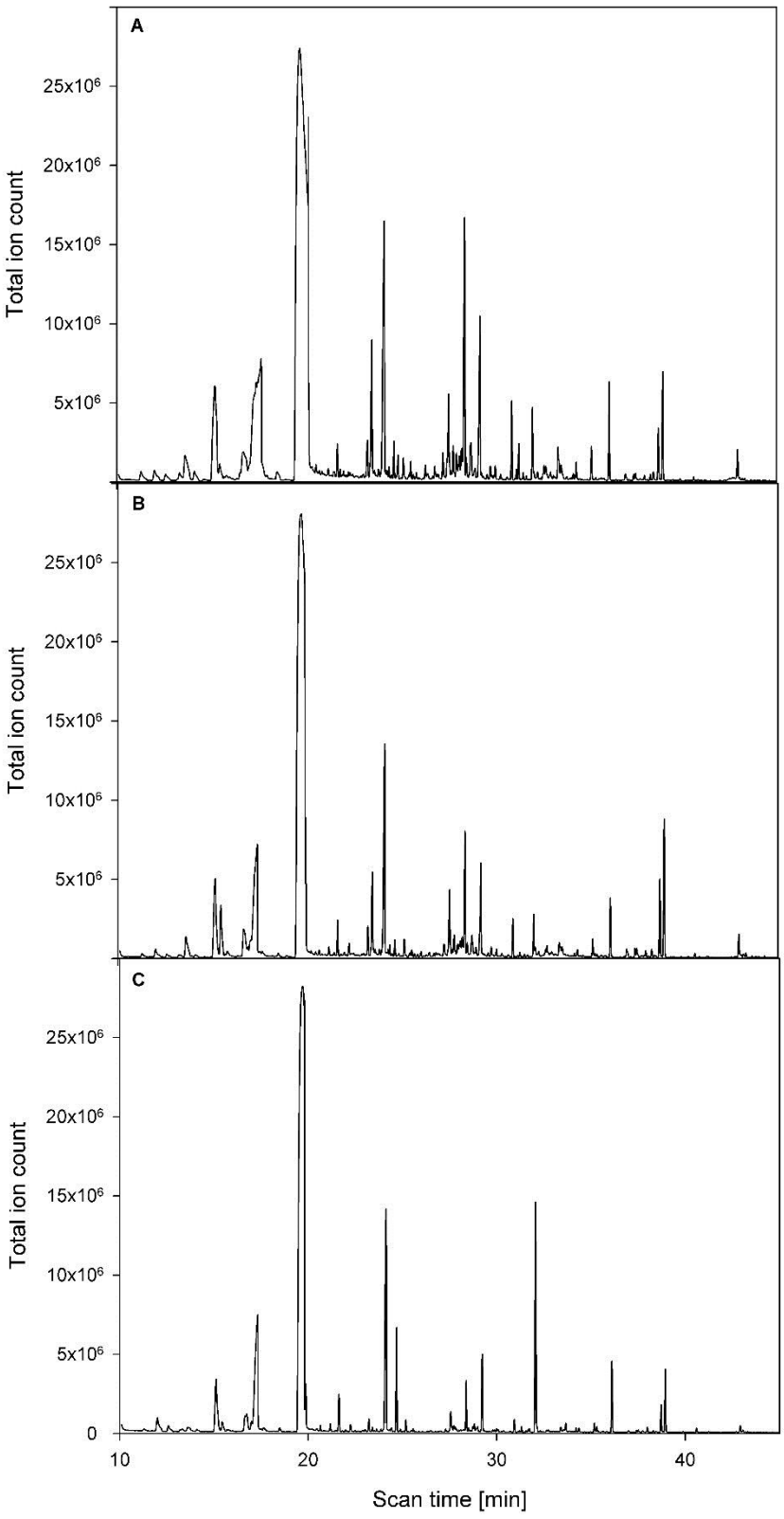
Aspartate



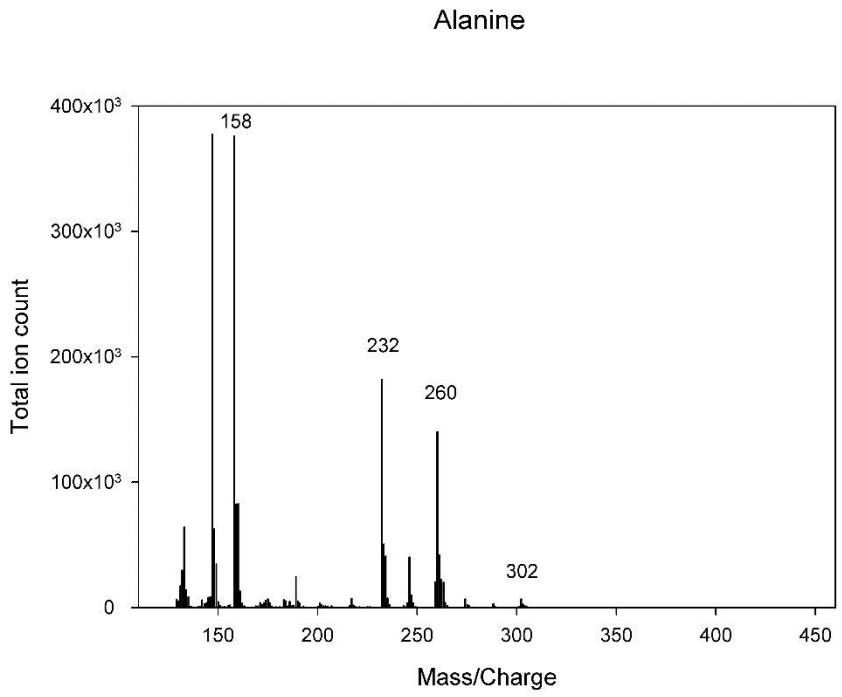
Glutamine

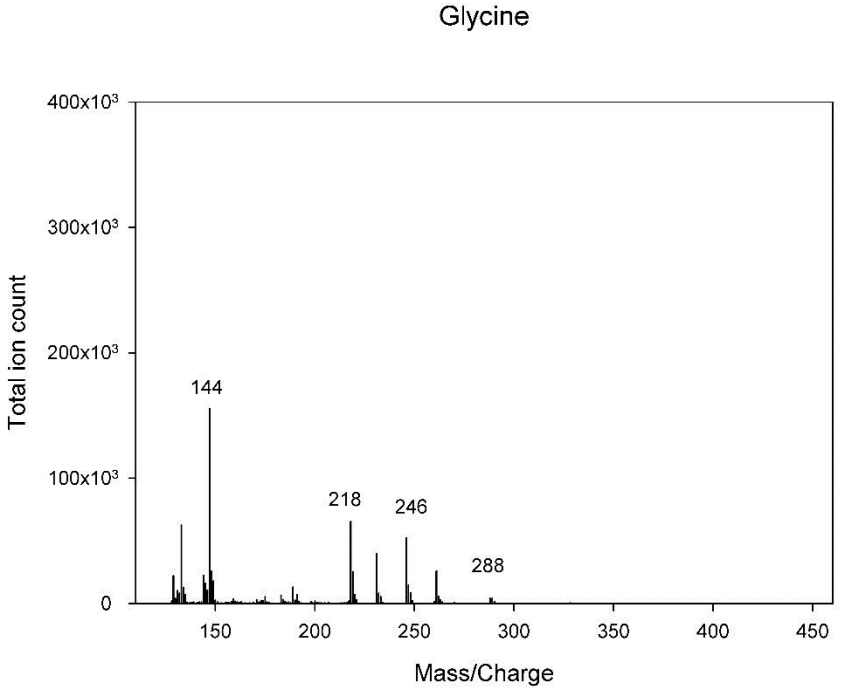


Appendix 9: **Chromatographic analysis of GFP samples processed using different volumes of derivatisation reagents.** GFP was immunopurified from one culture of 30-50 seedlings, hydrolysed and derivatised with a total volume of (A) 21 µl, (B) 43 µl or (C) 70 µl, of derivatisation reagents.

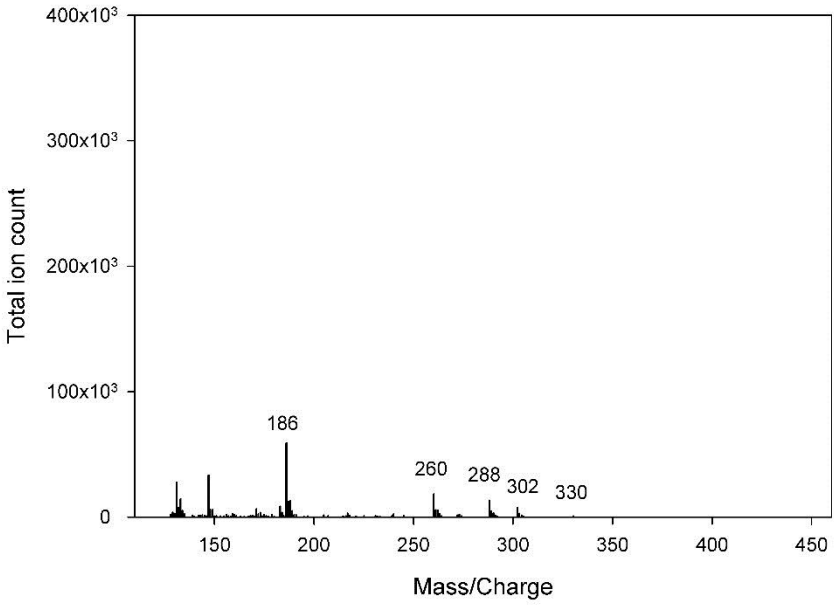


Appendix 10: **Mass spectra of hydrolysed GFP immunopurified from a single culture of 30-50 seedlings with constitutive GFP expression in reduced total volume of derivatisation reagents.**

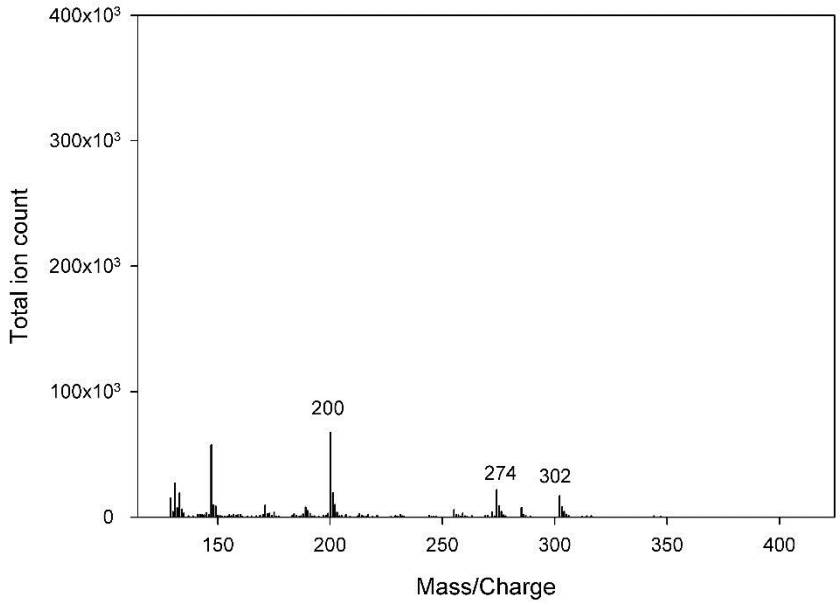




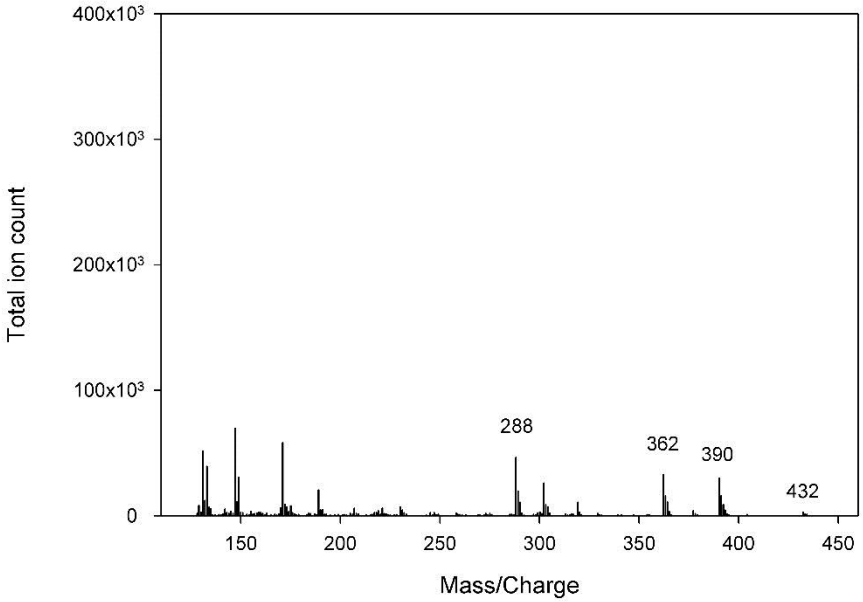
Valine



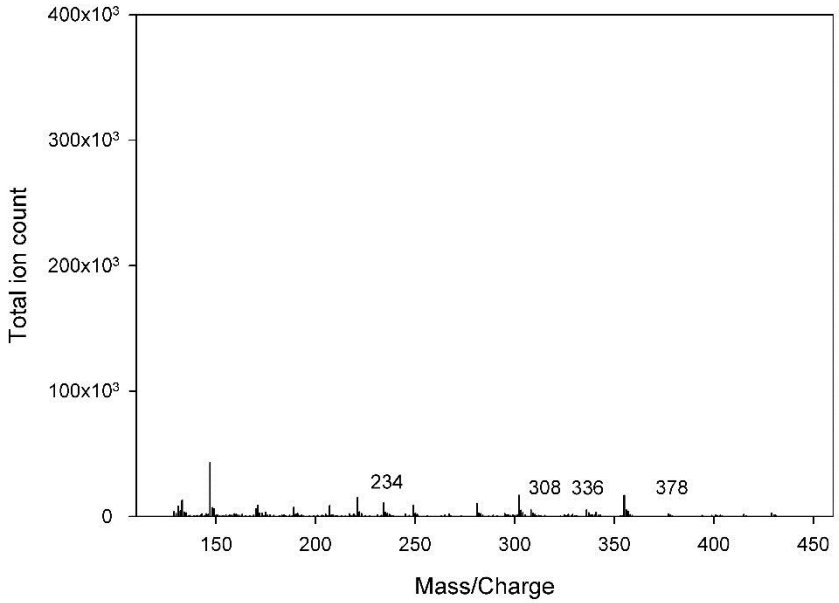
Leucine or isoleucine



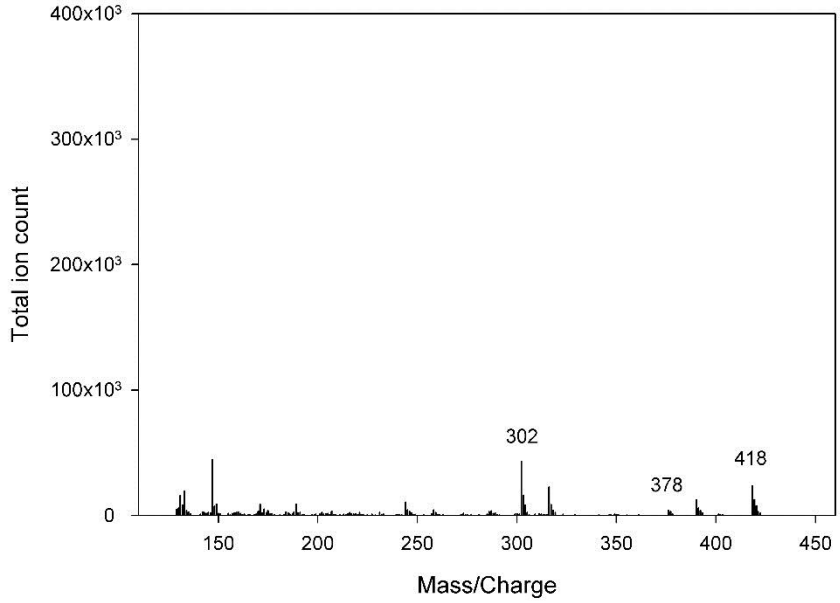
Serine



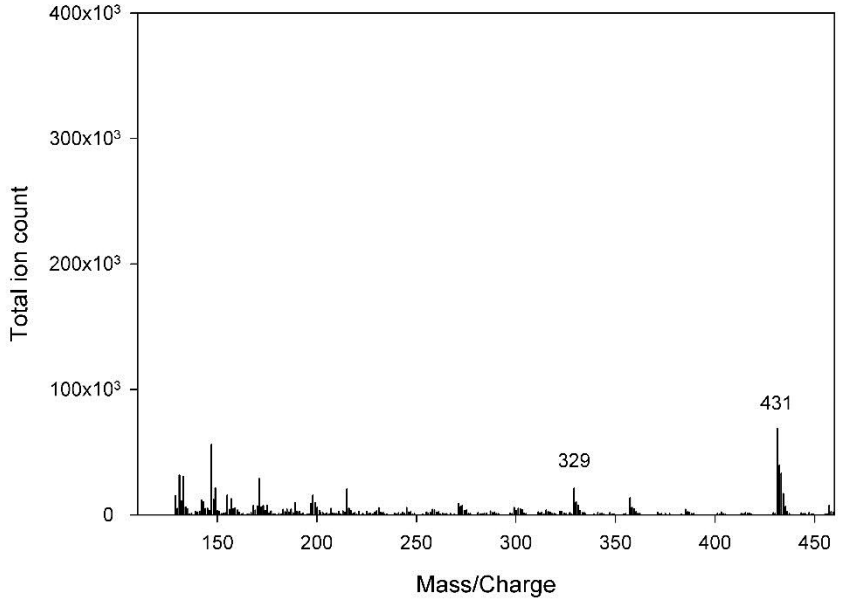
Phenylalanine



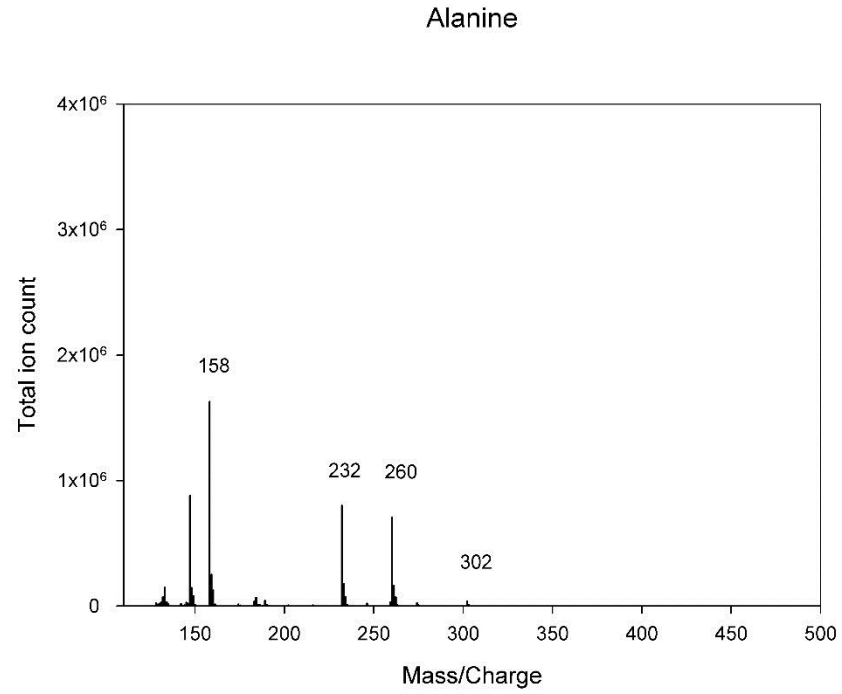
Aspartate

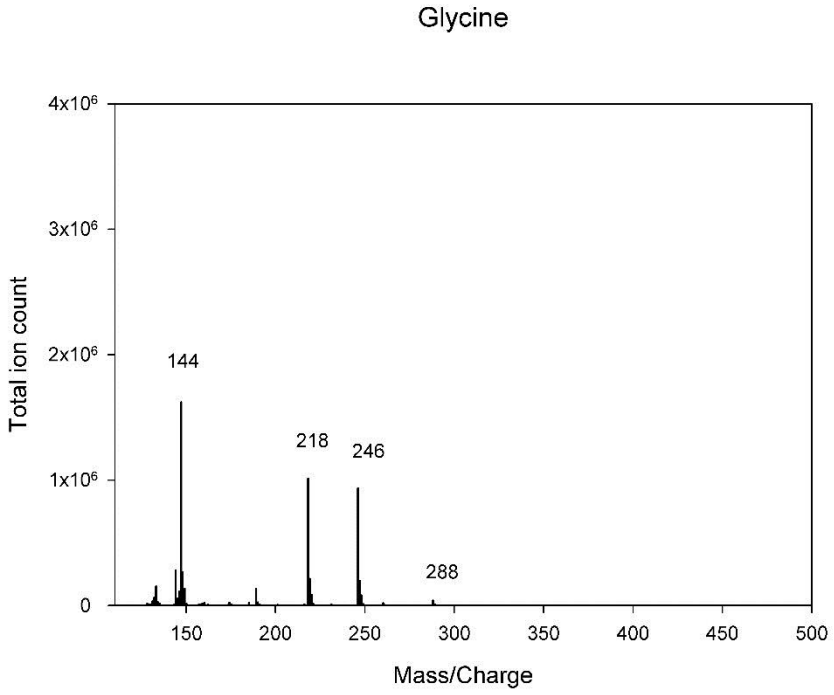


Glutamine

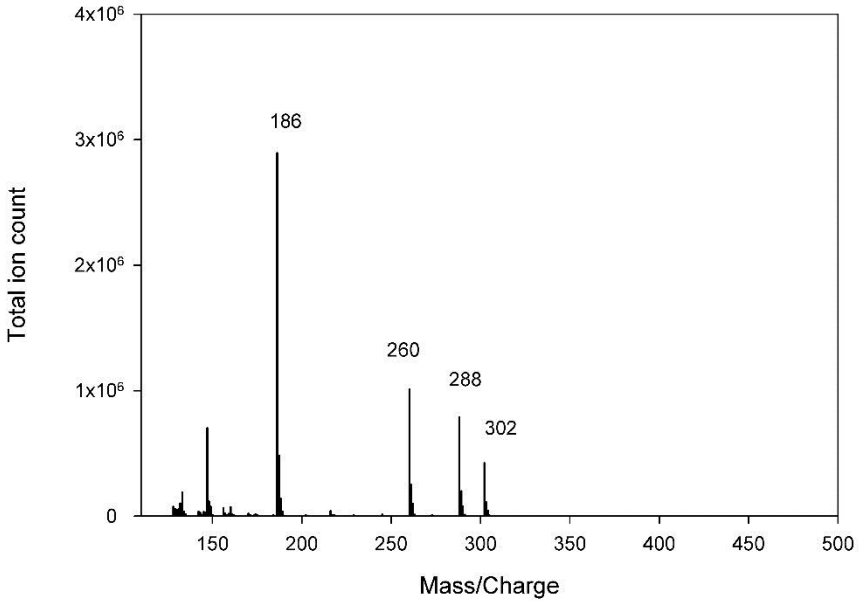


Appendix 11: **Mass spectra of amino acids derived from immunopurified GFP from three cultures of 30-50 seedlings hydrolysed and derivatised with a reduced total volume of derivatisation reagents (20 µl).**

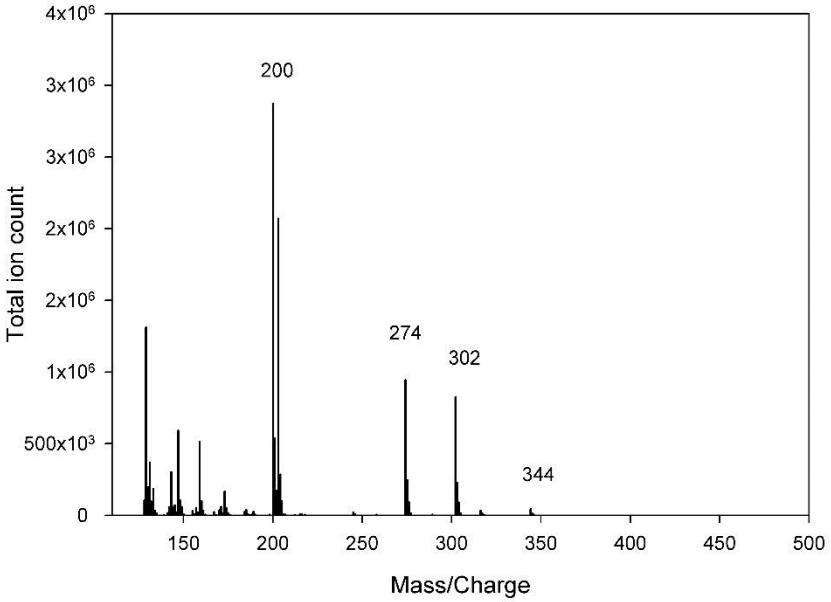




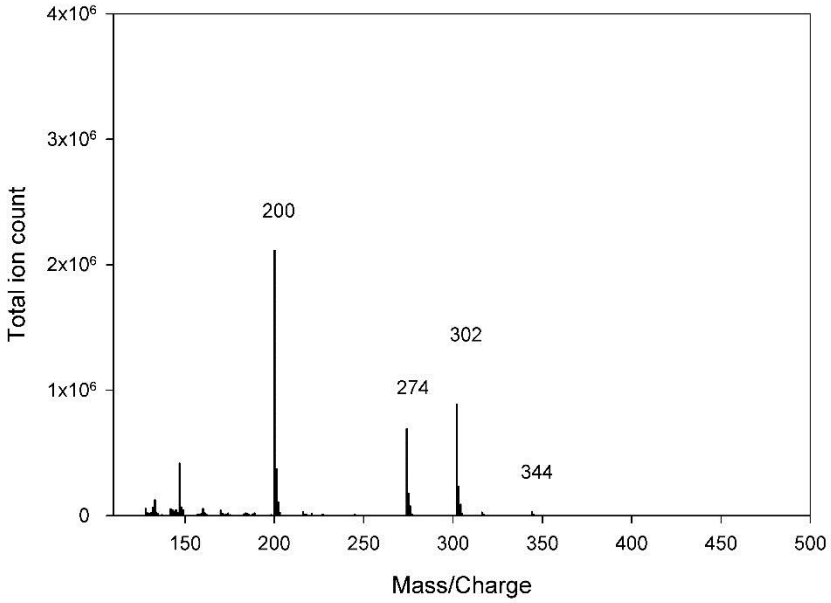
Valine



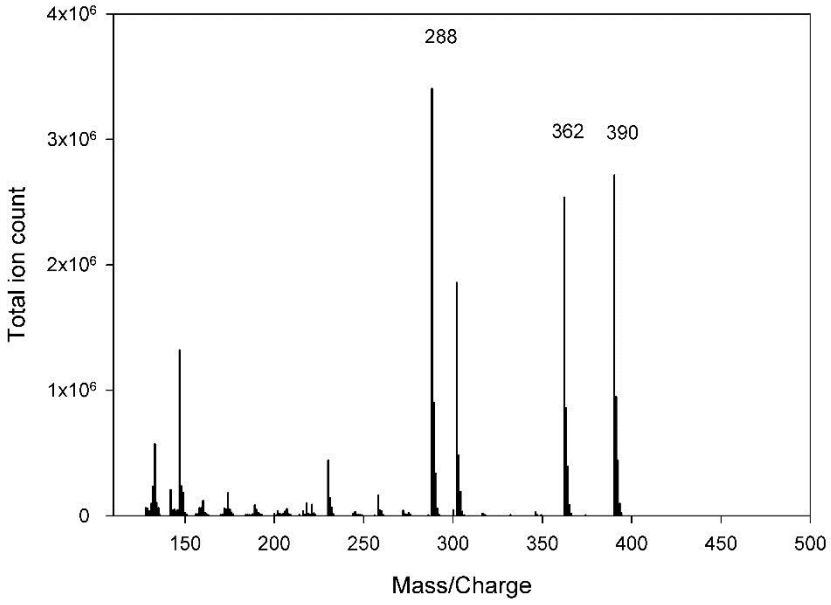
Leucine



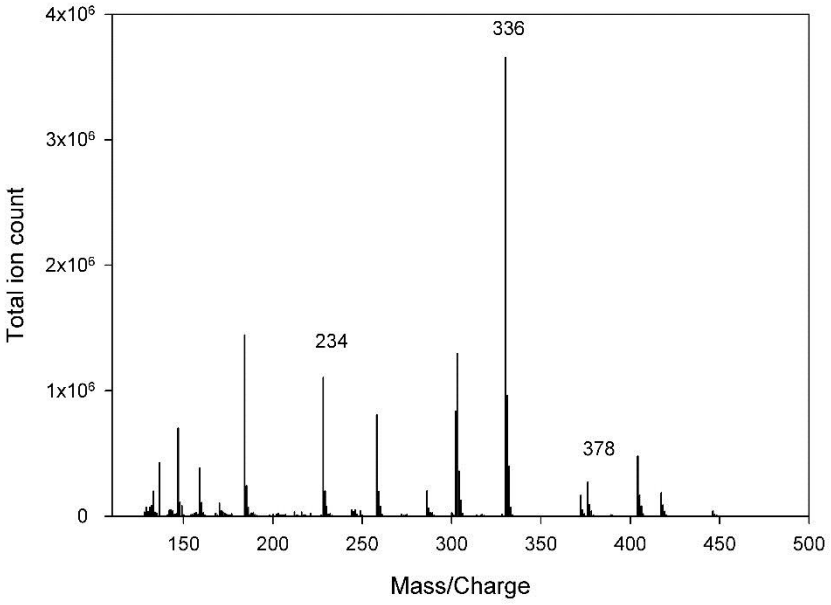
Isoleucine



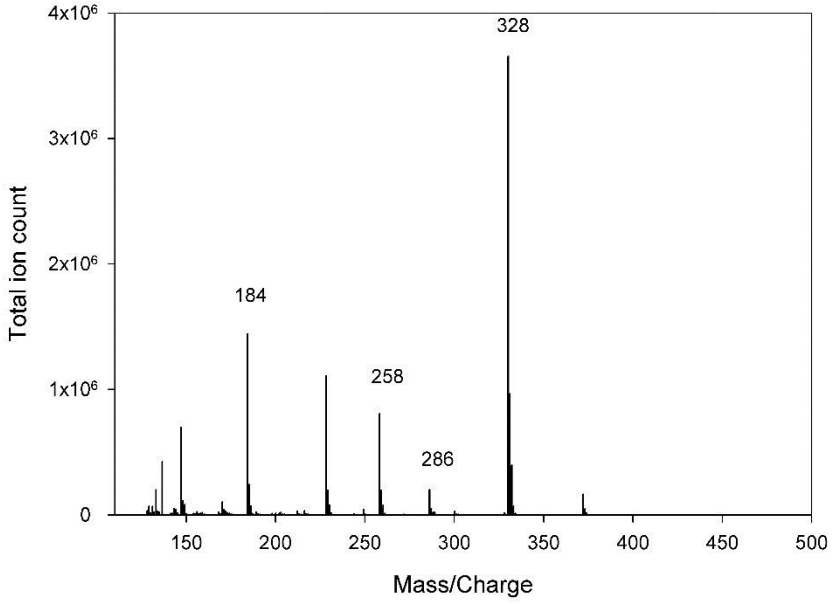
Serine



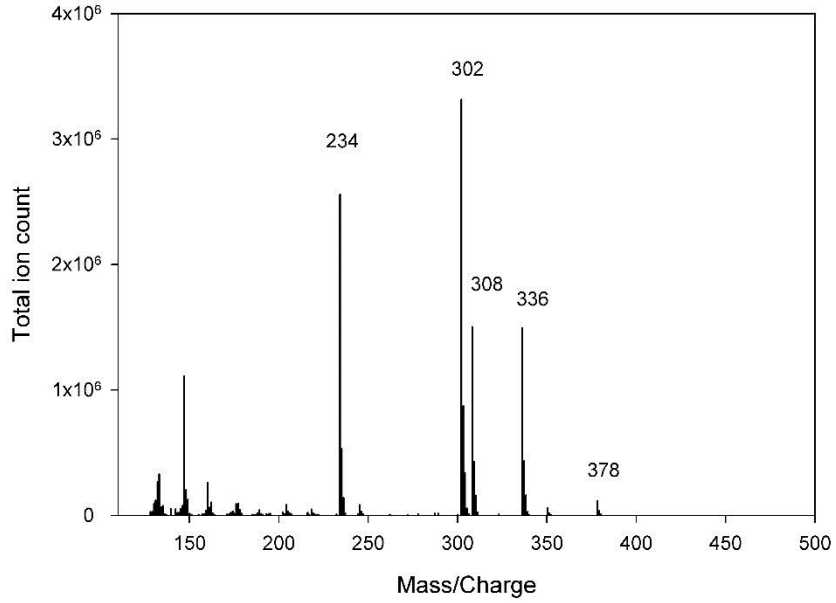
Threonine



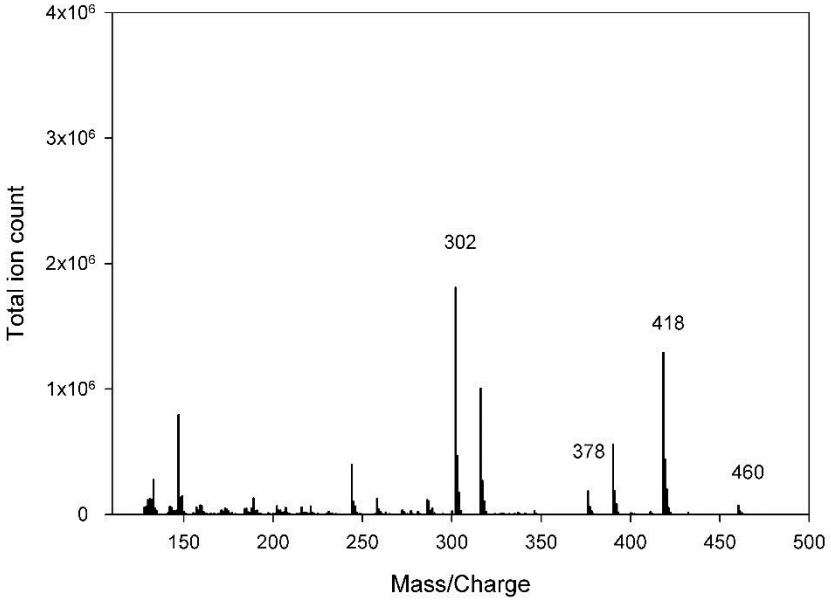
Proline



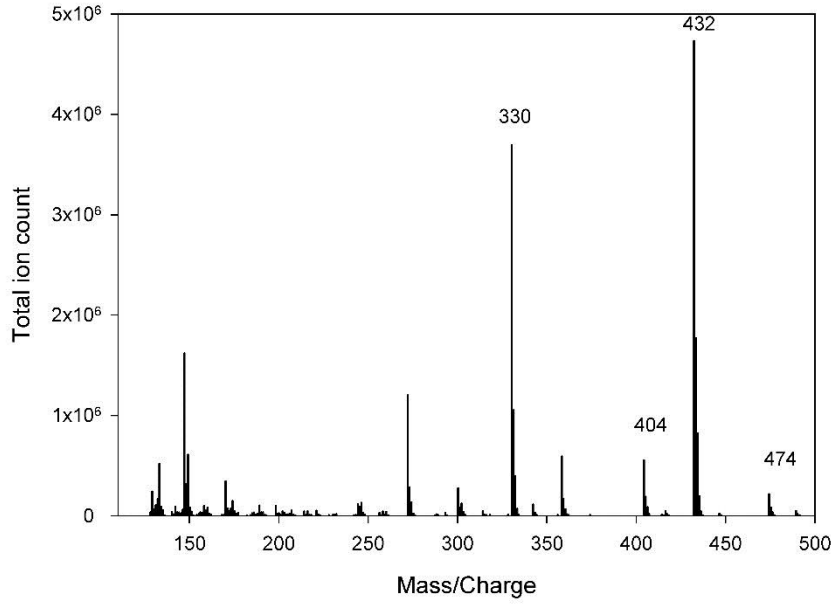
Phenylalanine



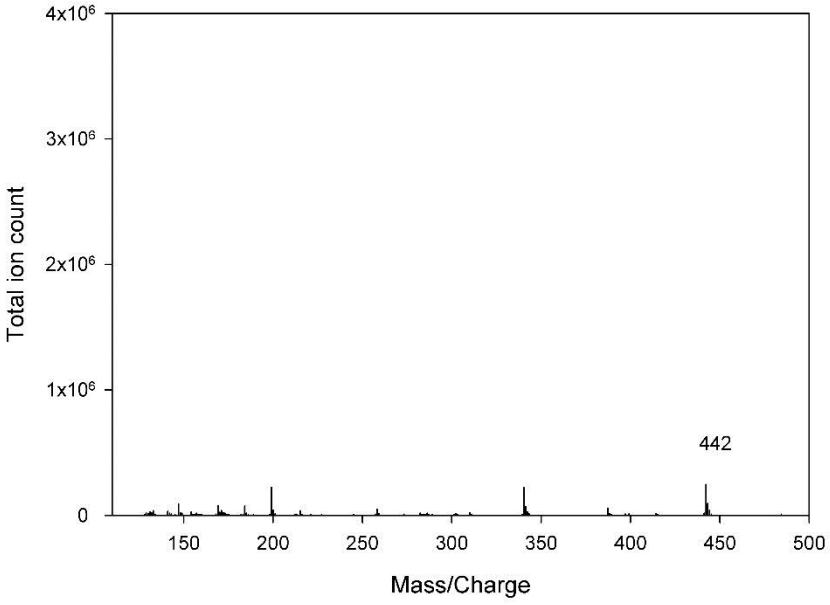
Aspartate



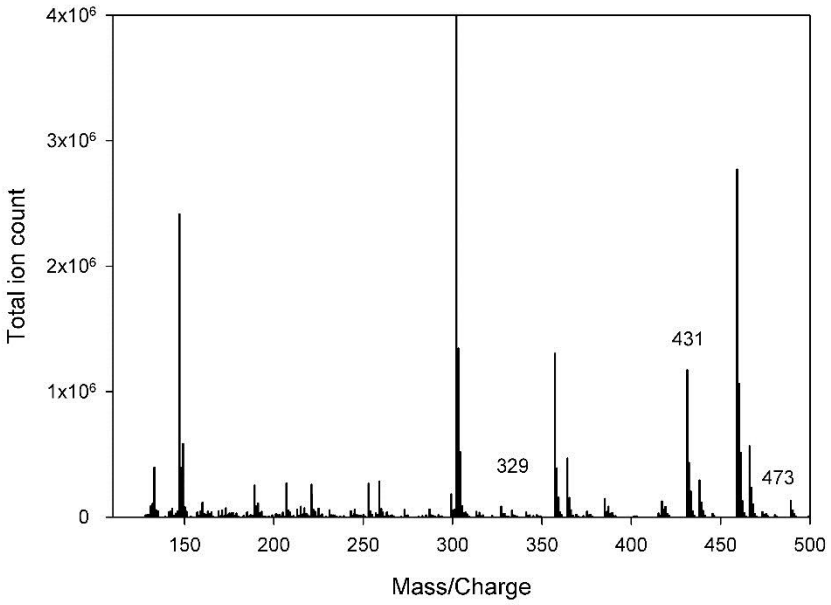
Glutamate



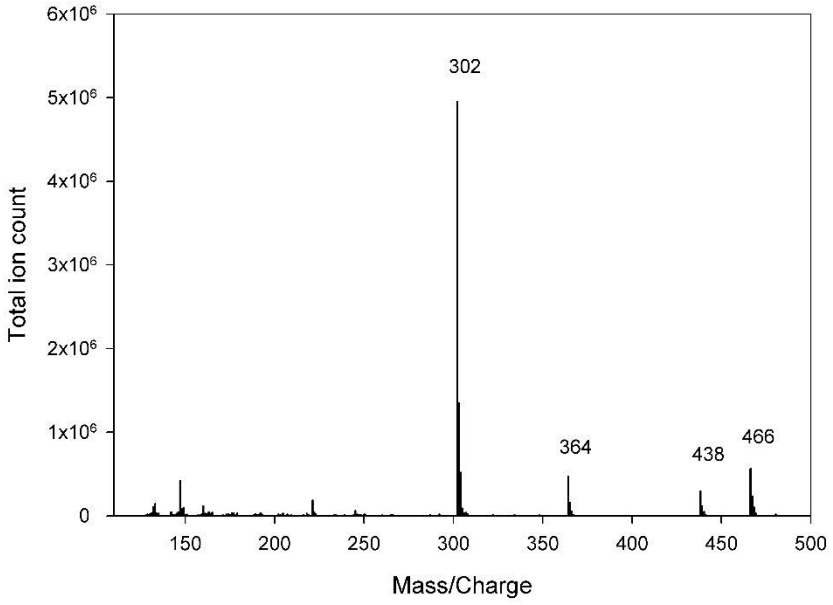
Arginine



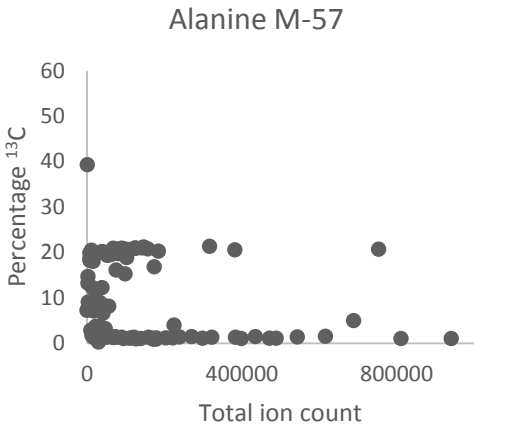
Lysine

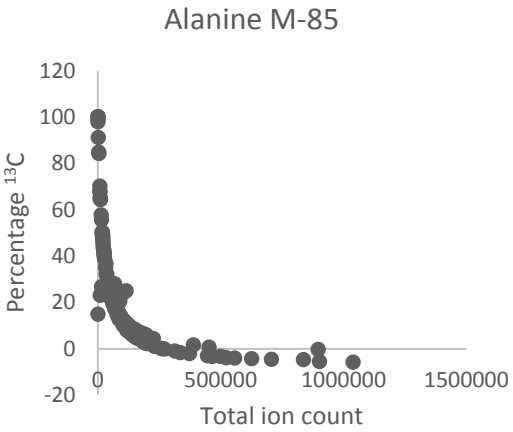


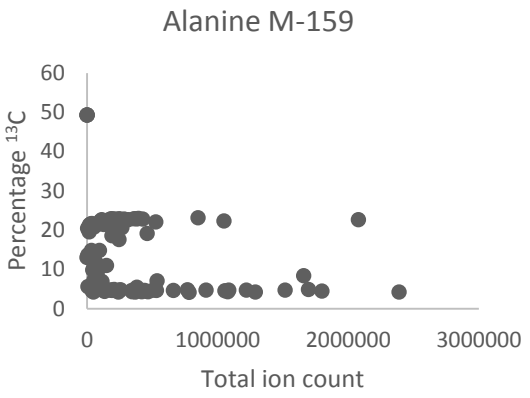
Tyrosine

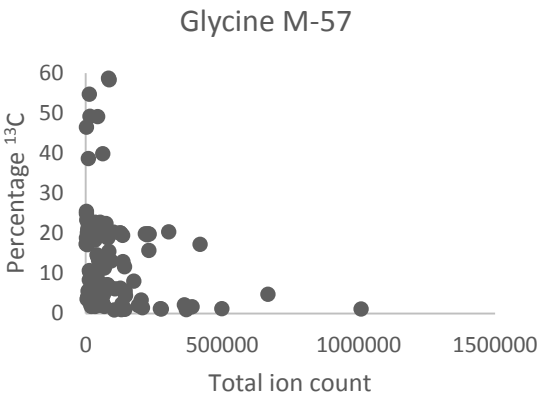


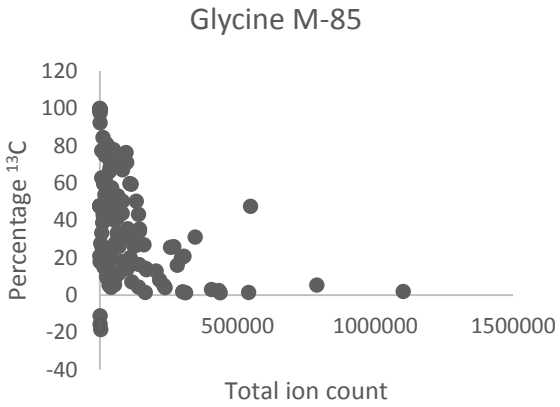
Appendix 12: **Comparison of total ion counts and fractional abundances for individual amino acid fragments detected in tBDMS derivatised protein hydrolysate.**

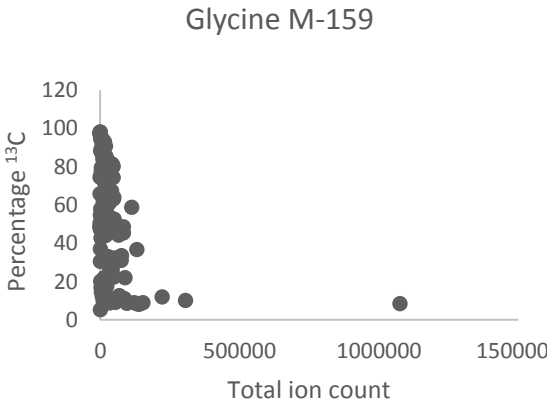


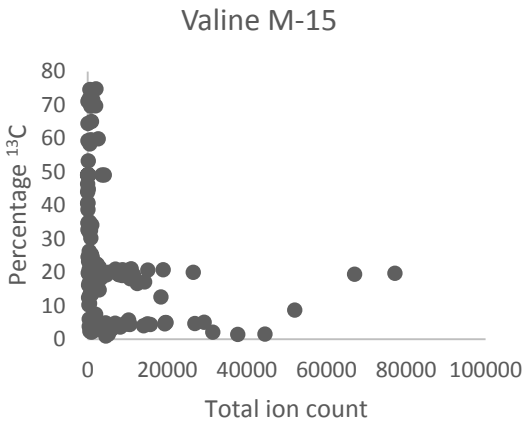


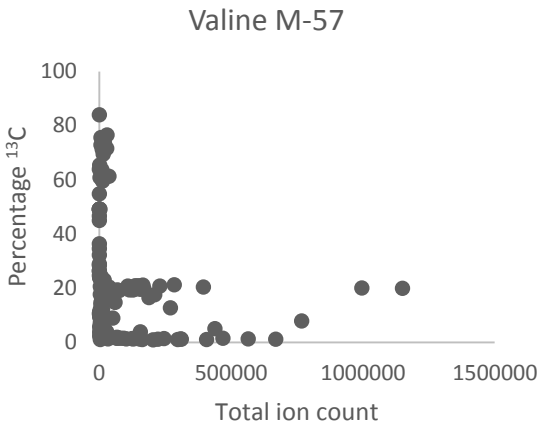


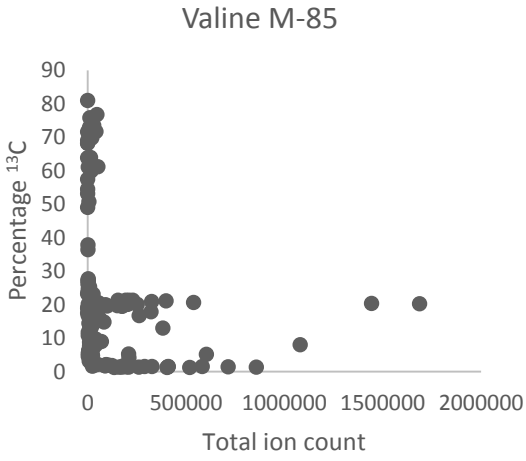


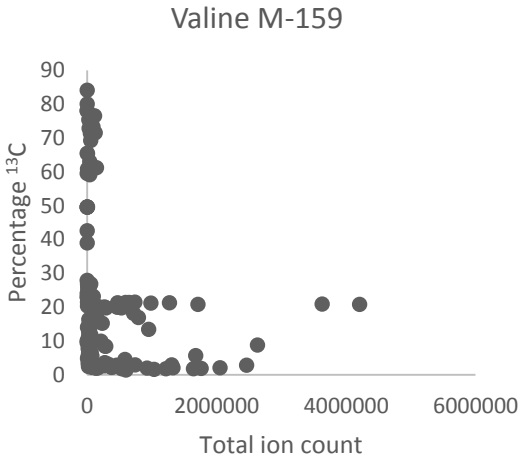


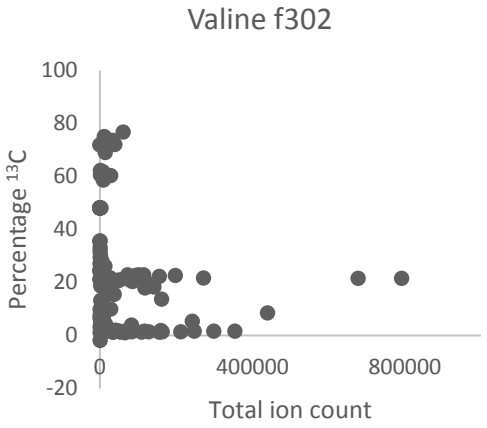


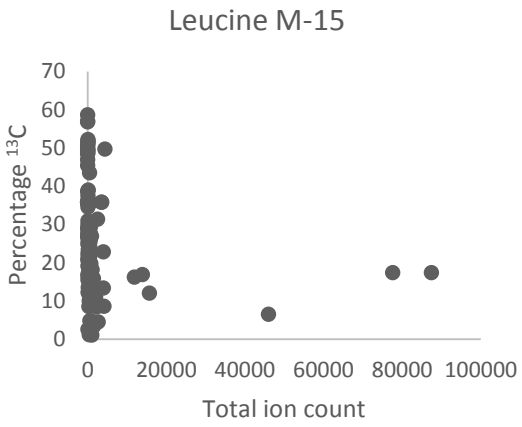

Valine M-15
Percentage 13C
Total ion count
0
10
20
30
40
50
60
70
80
0
20000
40000
60000
80000
100000

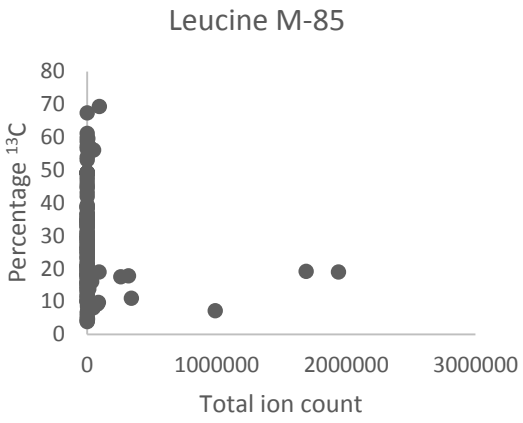

Valine M-57
Percentage 13C
Total ion count
0
20
40
60
80
100
0
500000
1000000
1500000

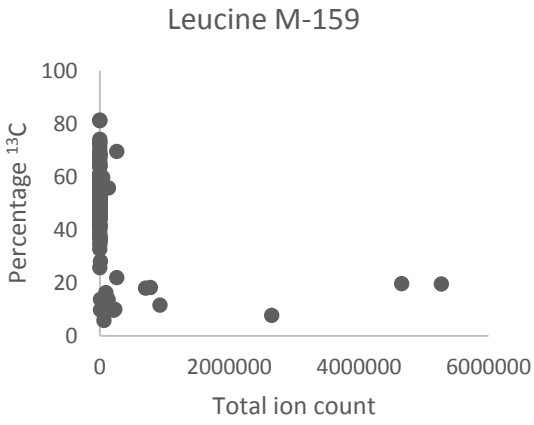

Valine M-85
Percentage 13C
Total ion count
0
10
20
30
40
50
60
70
80
90
0
500000
1000000
1500000
2000000

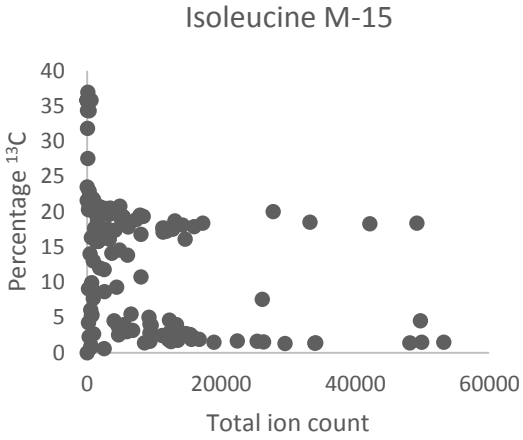

Valine M-159
Percentage 13C
Total ion count
0
10
20
30
40
50
60
70
80
90
0
2000000
4000000
6000000

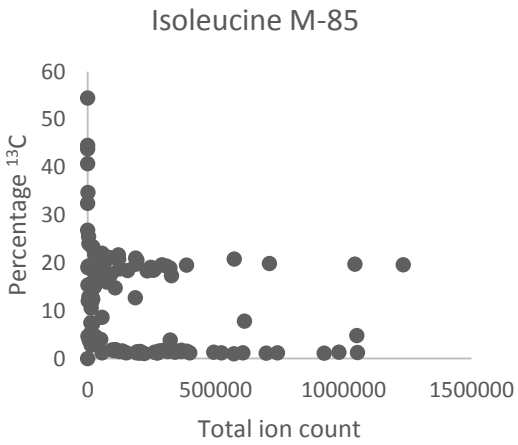

Valine f302
Percentage 13C
Total ion count
-20
0
20
40
60
80
100
0
400000
800000

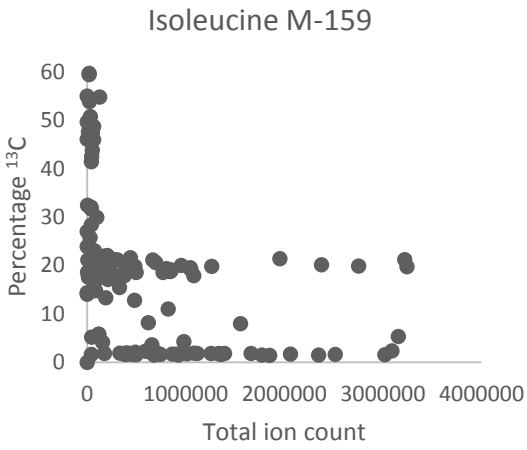

Leucine M-15
Percentage 13C
70
60
50
40
30
20
10
0
0
20000
40000
60000
80000
100000
Total ion count

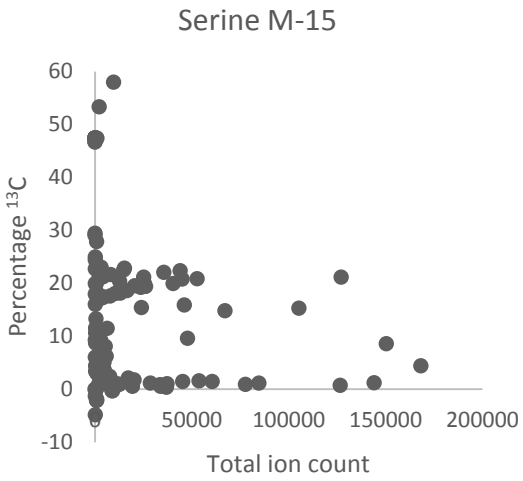

Leucine M-85
Percentage 13C
80
70
60
50
40
30
20
10
0
0
1000000
2000000
3000000
Total ion count

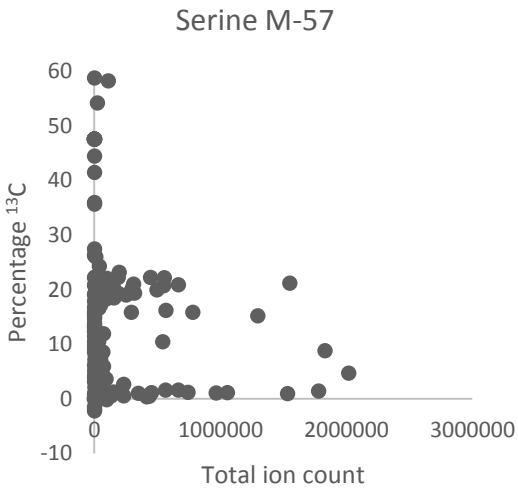

Leucine M-159
Percentage 13C
100
80
60
40
20
0
0
2000000
4000000
6000000
Total ion count

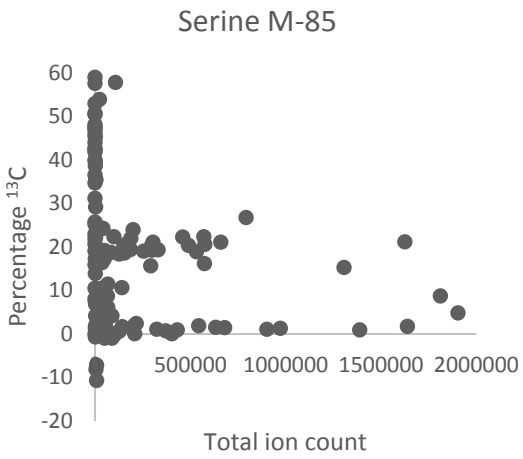

Isoleucine M-15
Percentage 13C
40
35
30
25
20
15
10
5
0
0
20000
40000
60000
Total ion count

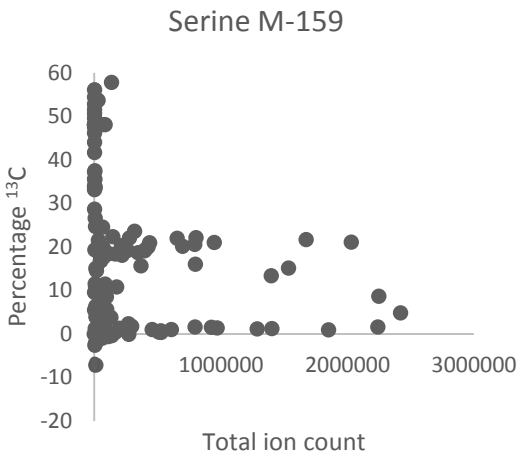

Isoleucine M-85
Percentage 13C
60
50
40
30
20
10
0
0
500000
1000000
1500000
Total ion count

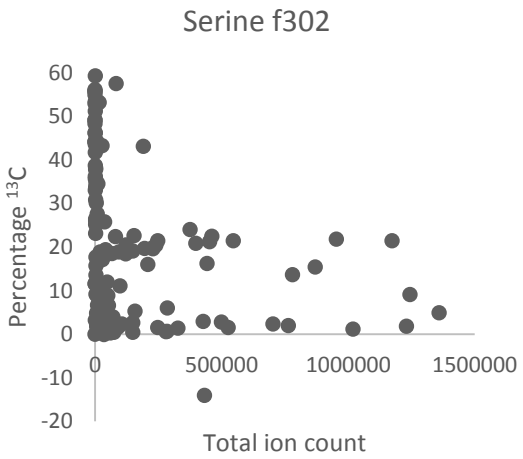

Isoleucine M-159
Percentage 13C
60
50
40
30
20
10
0
0
1000000
2000000
3000000
4000000
Total ion count

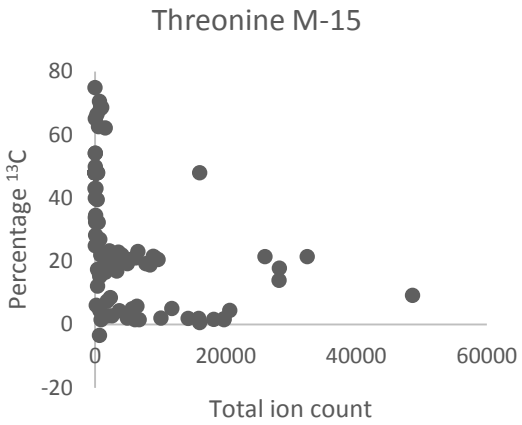

Serine M-15
Percentage 13C
60
50
40
30
20
10
0
-10
0
50000
100000
150000
200000
Total ion count

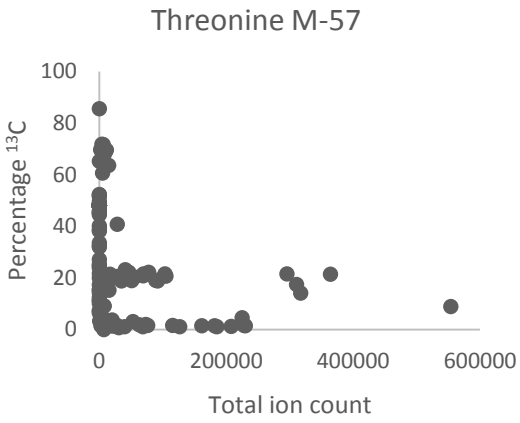

Serine M-57
Percentage 13C
60
50
40
30
20
10
0
-10
0
1000000
2000000
3000000
Total ion count

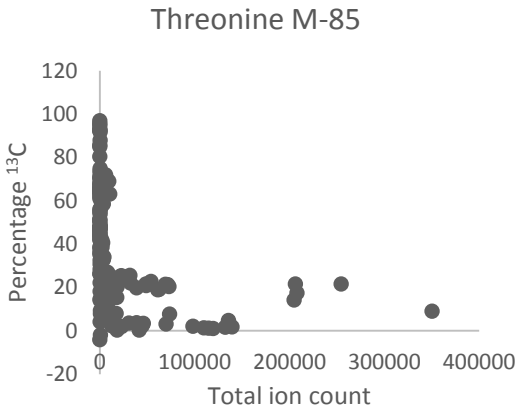

Serine M-85
Percentage 13C
60
50
40
30
20
10
0
-10
-20
500000
1000000
1500000
2000000
Total ion count

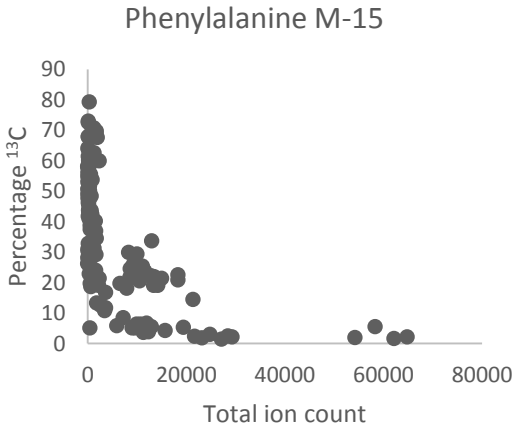

Serine M-159
Percentage 13C
60
50
40
30
20
10
0
-10
-20
1000000
2000000
3000000
Total ion count

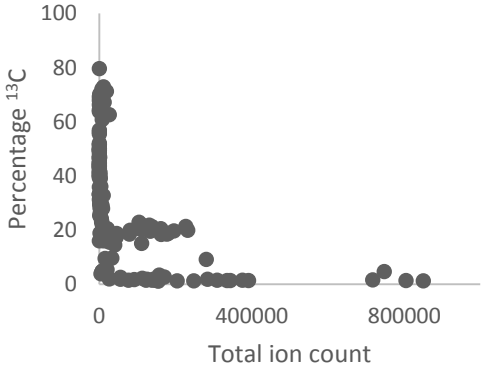

Serine f302
Percentage 13C
60
50
40
30
20
10
0
-10
-20
0
500000
1000000
1500000
Total ion count

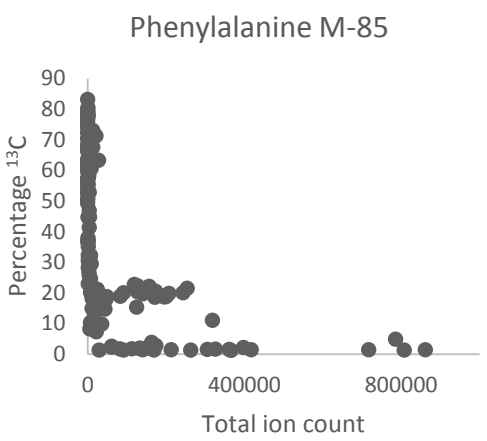

Threonine M-15
Percentage ^{13}C
80
60
40
20
0
-20
0
20000
40000
60000
Total ion count

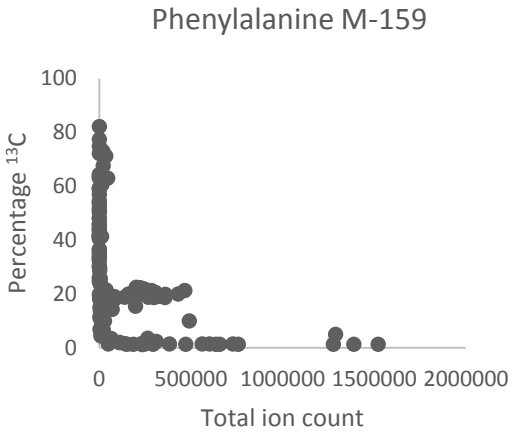

Threonine M-57
Percentage ^{13}C
100
80
60
40
20
0
0
200000
400000
600000
Total ion count

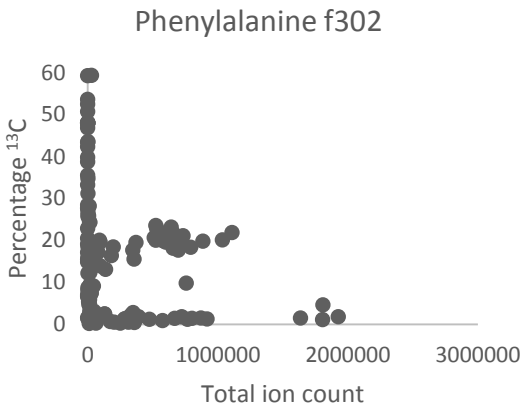

Threonine M-85
Percentage ^{13}C
120
100
80
60
40
20
0
-20
0
100000
200000
300000
400000
Total ion count

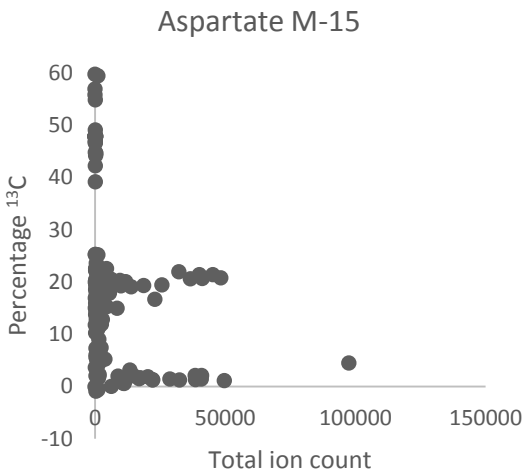

Phenylalanine M-15
Percentage ^{13}C
90
80
70
60
50
40
30
20
10
0
0
20000
40000
60000
80000
Total ion count

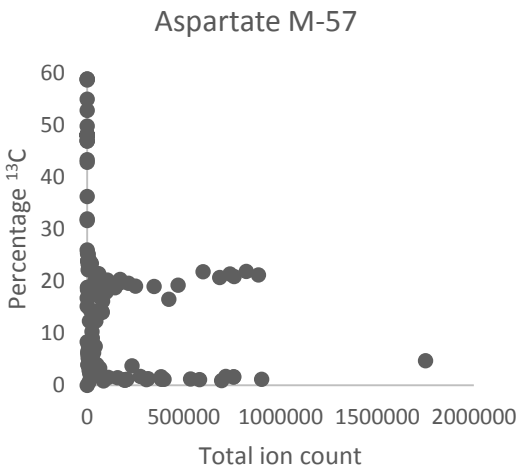

Phenylalanine M-57
Percentage ^{13}C
100
80
60
40
20
0
0
400000
800000
Total ion count

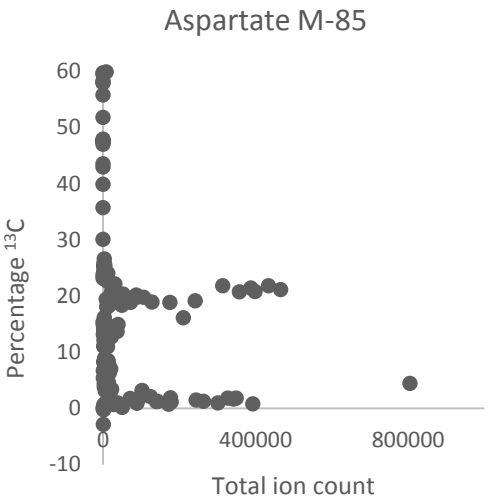

Phenylalanine M-85
Percentage 13C
90
80
70
60
50
40
30
20
10
0
0
400000
800000
Total ion count

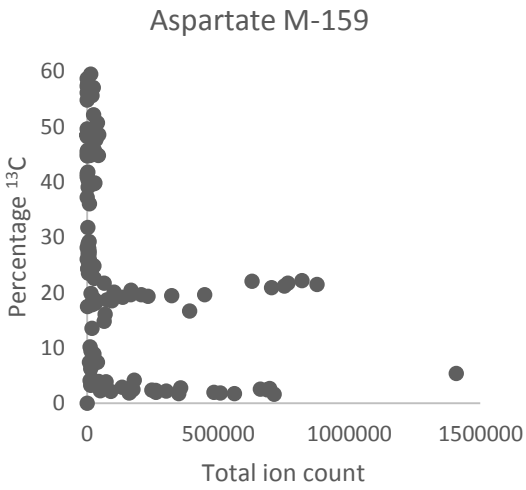

Phenylalanine M-159
Percentage 13C
100
80
60
40
20
0
0
500000
1000000
1500000
2000000
Total ion count

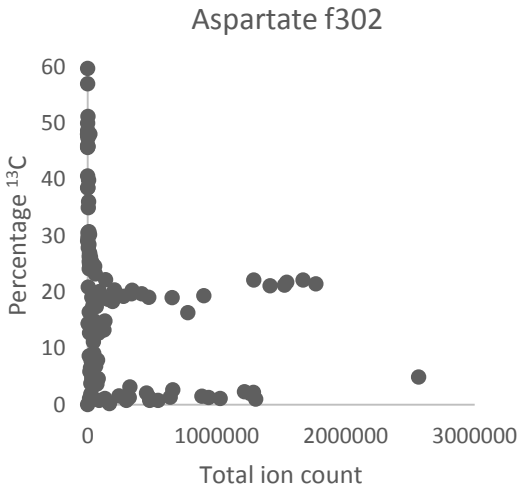

Phenylalanine f302
Percentage 13C
60
50
40
30
20
10
0
0
1000000
2000000
3000000
Total ion count

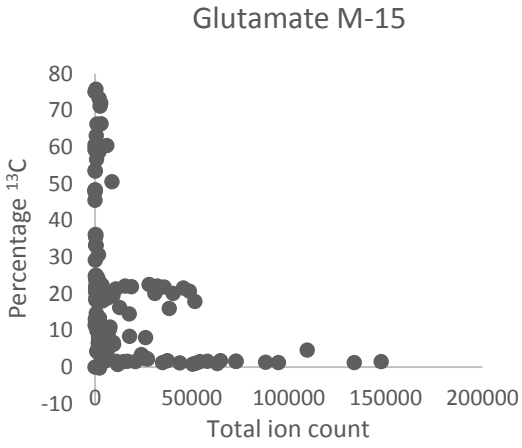

Aspartate M-15
Percentage 13C
60
50
40
30
20
10
0
-10
0
50000
100000
150000
Total ion count

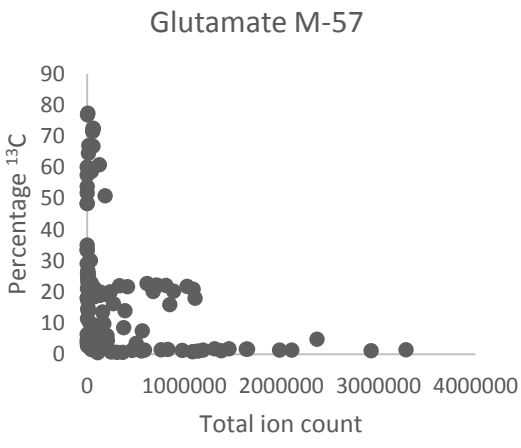

Aspartate M-57
Percentage 13C
60
50
40
30
20
10
0
0
500000
1000000
1500000
2000000
Total ion count

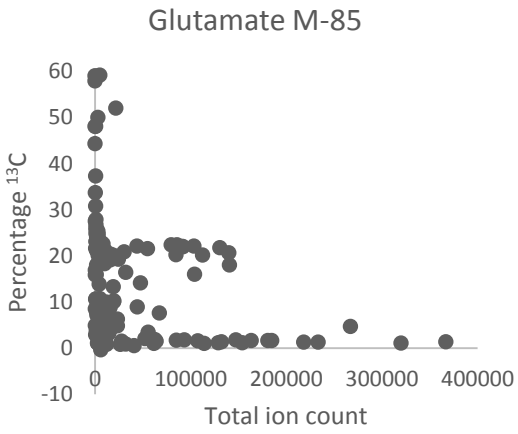

Aspartate M-85
Percentage 13C
60
50
40
30
20
10
0
-10
0
400000
800000
Total ion count

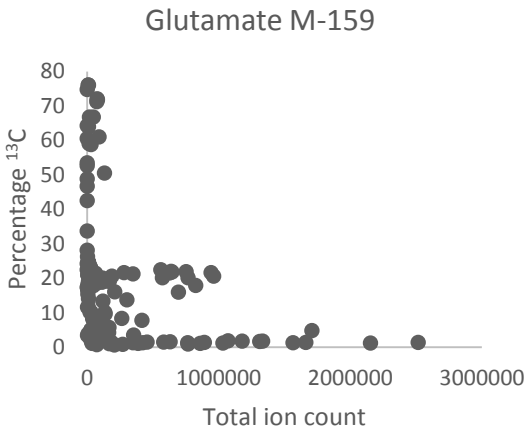

Aspartate M-159
Percentage 13C
60
50
40
30
20
10
0
0
500000
1000000
1500000
Total ion count

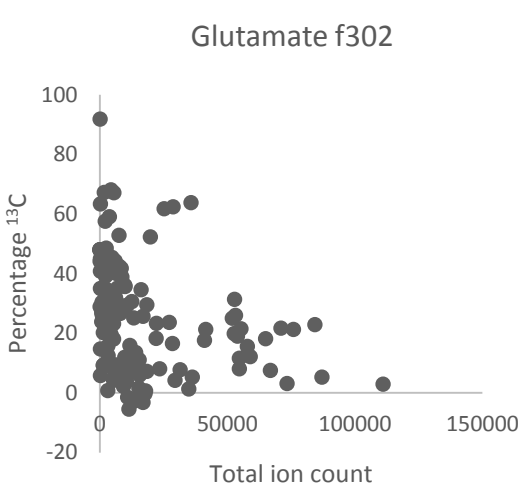

Aspartate f302
Percentage 13C
60
50
40
30
20
10
0
0
1000000
2000000
3000000
Total ion count


Glutamate M-15
Percentage 13C
80
70
60
50
40
30
20
10
0
-10
0
50000
100000
150000
200000
Total ion count

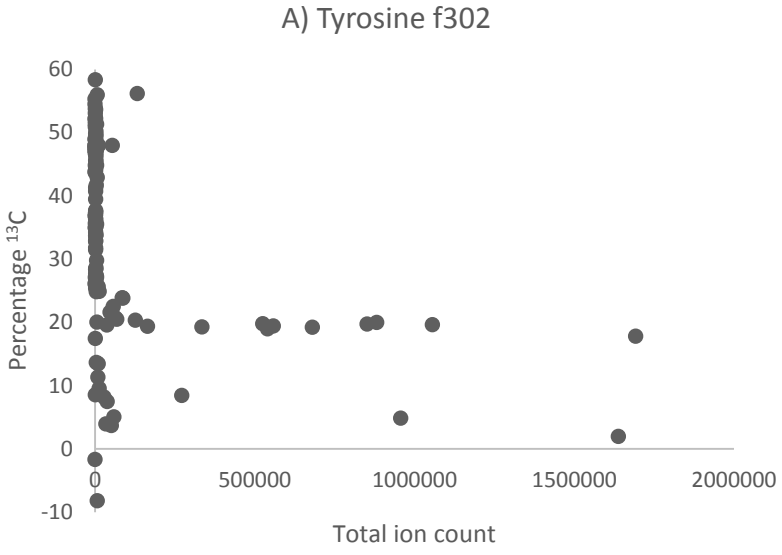

Glutamate M-57
Percentage 13C
90
80
70
60
50
40
30
20
10
0
0
1000000
2000000
3000000
4000000
Total ion count

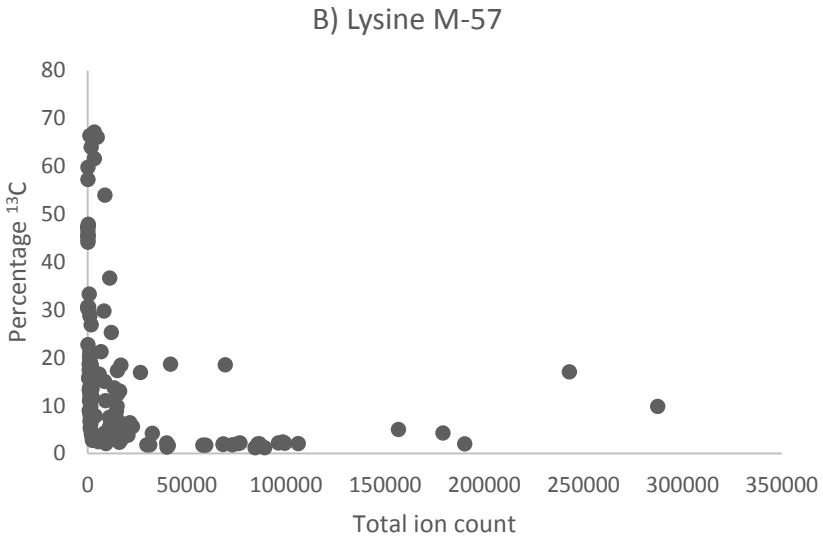

Glutamate M-85
Percentage 13C
60
50
40
30
20
10
0
-10
0
100000
200000
300000
400000
Total ion count


Glutamate M-159
Percentage 13C
80
70
60
50
40
30
20
10
0
0
1000000
2000000
3000000
Total ion count

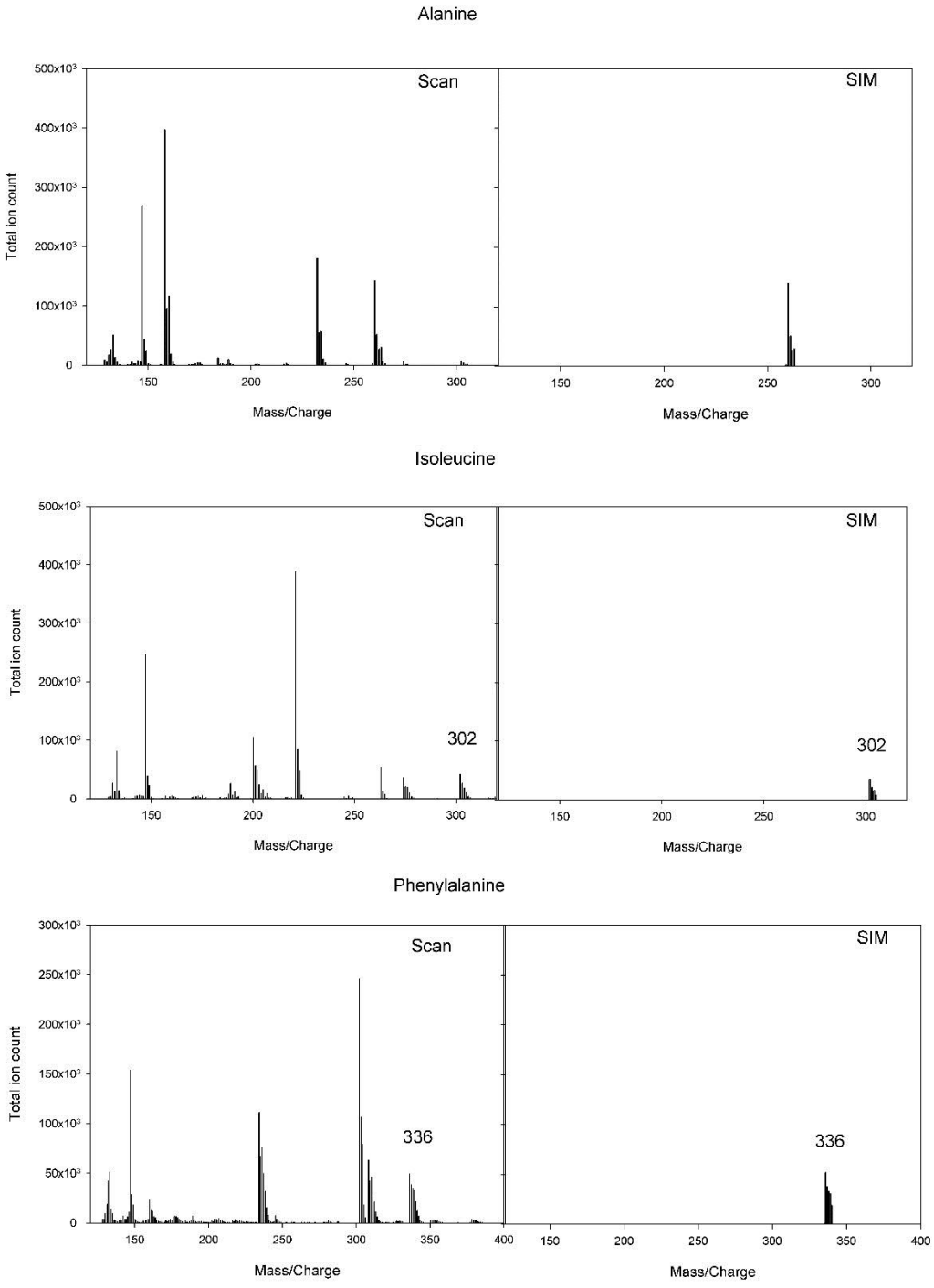

Glutamate f302
Percentage 13C
100
80
60
40
20
0
-20
0
50000
100000
150000
Total ion count

Appendix 13: **Examples of fragments not present at consistently high enough levels for analysis and not included in the data collated in Table 6:** (A) Tyrosine f302 fragment with only a few higher abundance measurements for samples with low levels (natural abundance) of ^{13}C enrichment. (B) Lysine M-57 fragment with only a few higher abundance measurements (20%).

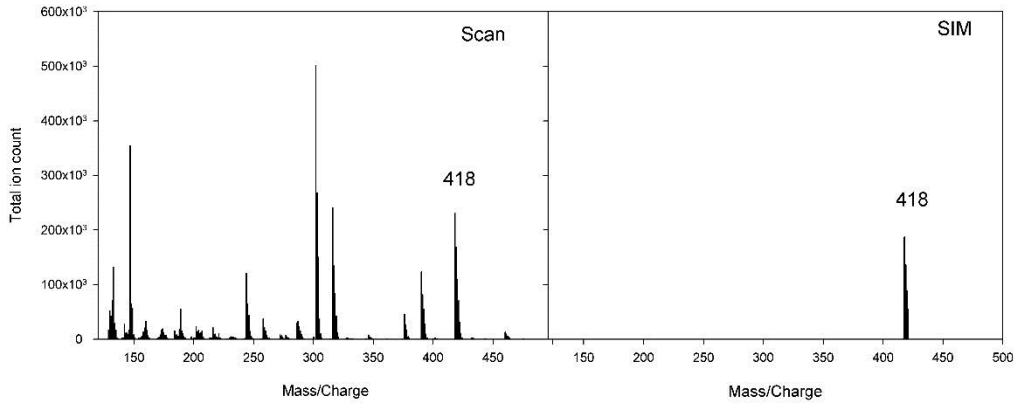




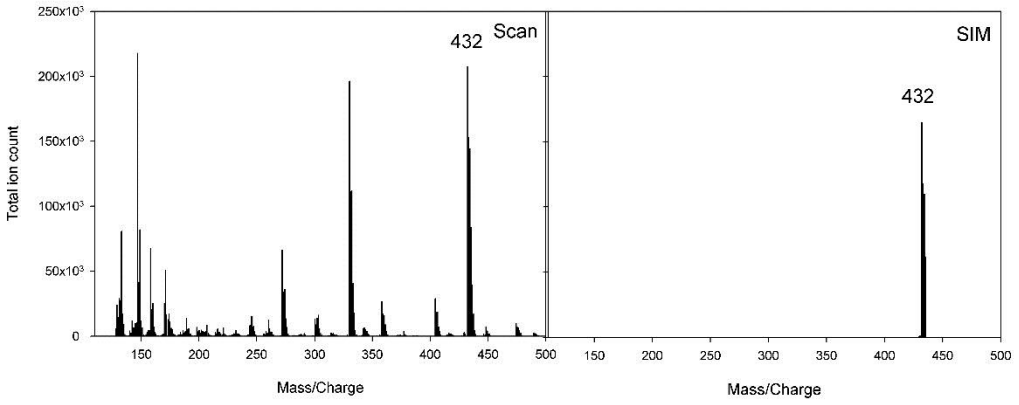
Appendix 14 : **Mass spectra of hydrolysed GFP immunopurified from a single culture of 30-50 seedlings with constitutive GFP expression and derivatised in a reduced volume of reagents.** Pairs of panels present SIM and scan runs of the same sample



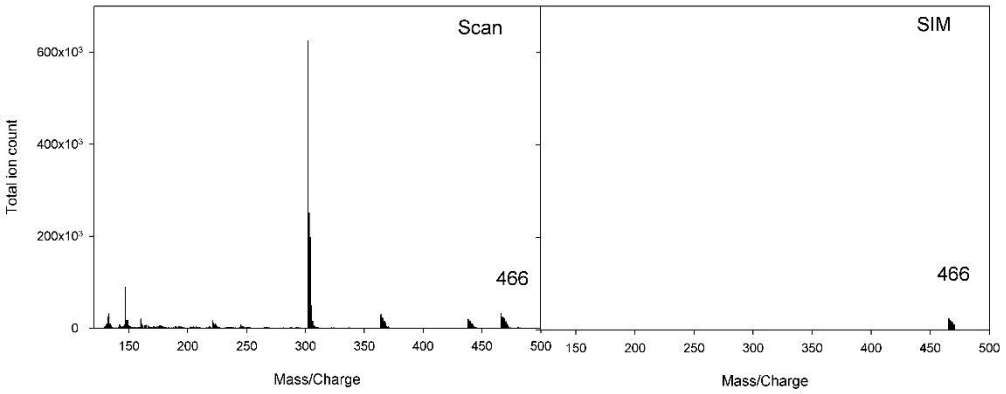
Aspartate



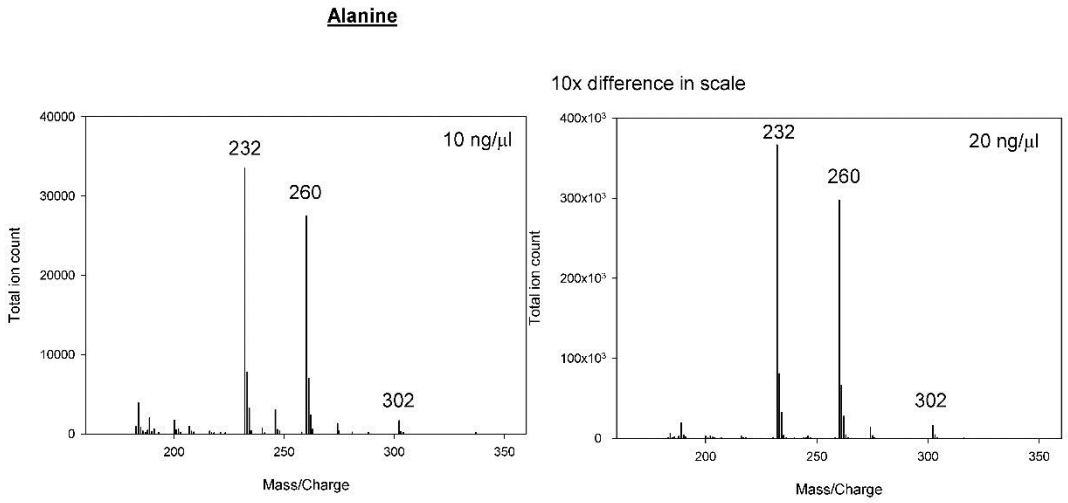
Glutamate

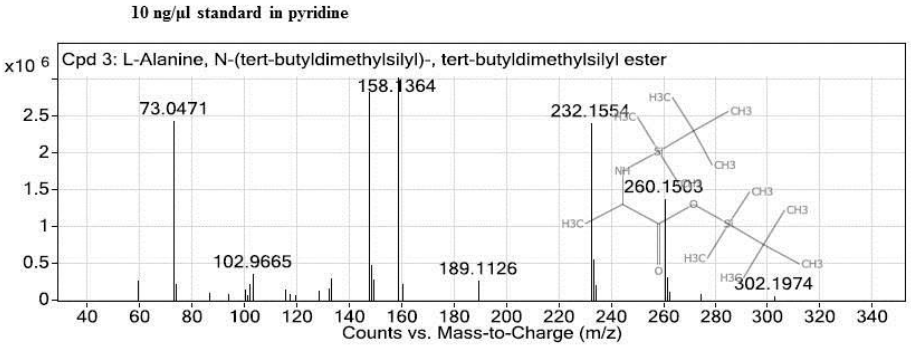


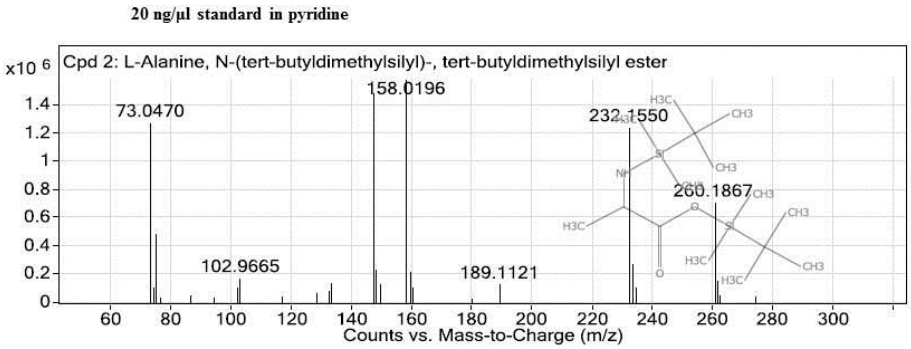
Tyrosine



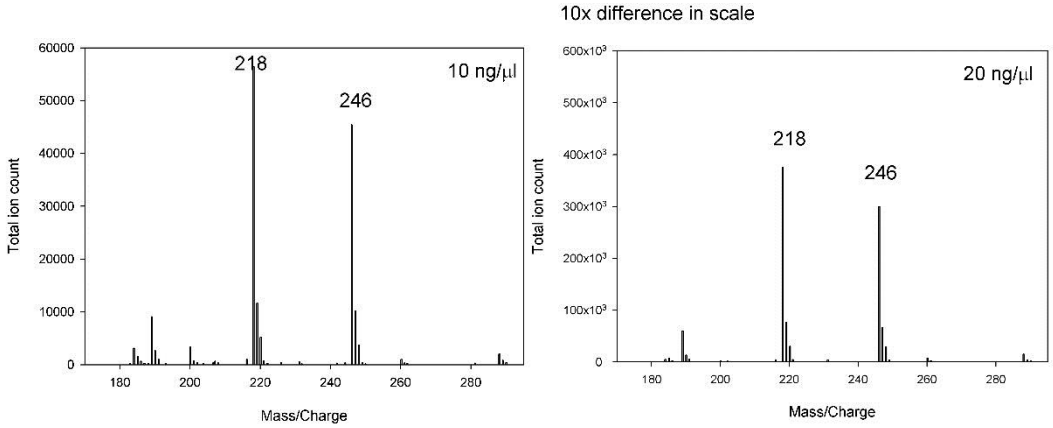
Appendix 15: **Comparison of mass spectra of amino acid fragments detected using Agilent 5975 and 7200 systems.** Individual amino acids were analysed using the same two protein standards in both systems. The amino acid data measured with Agilent 5975 system was extracted from the spectra and the data obtained with 7200 system is presented as summarised by an automated reporting system utilising the analysis software (Agilent Mass Hunter). Isoleucine and proline were not detected in one of the lower concentration samples.



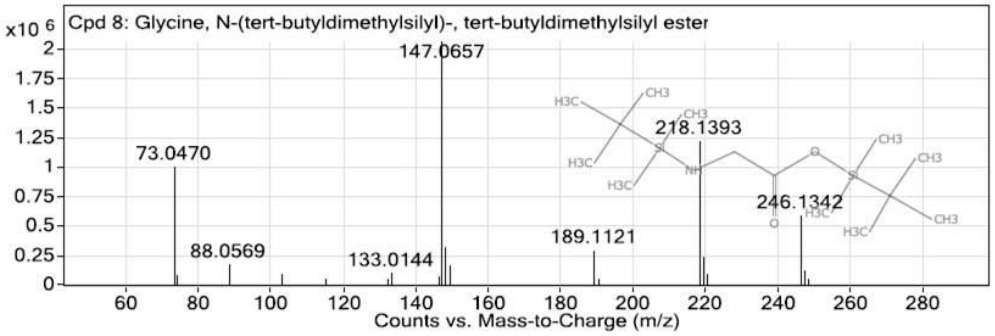




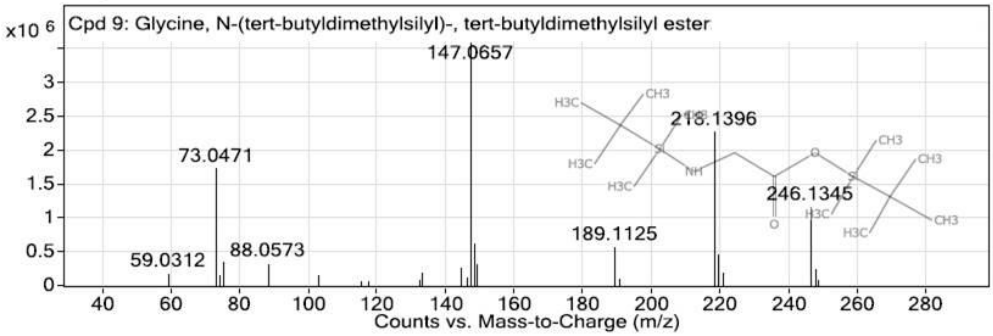
Glycine



10 ng/µl standard in pyridine

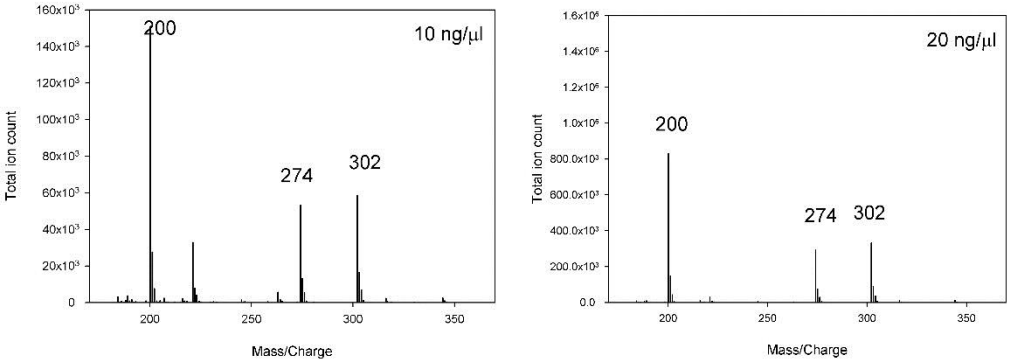


20 ng/µl standard in pyridine

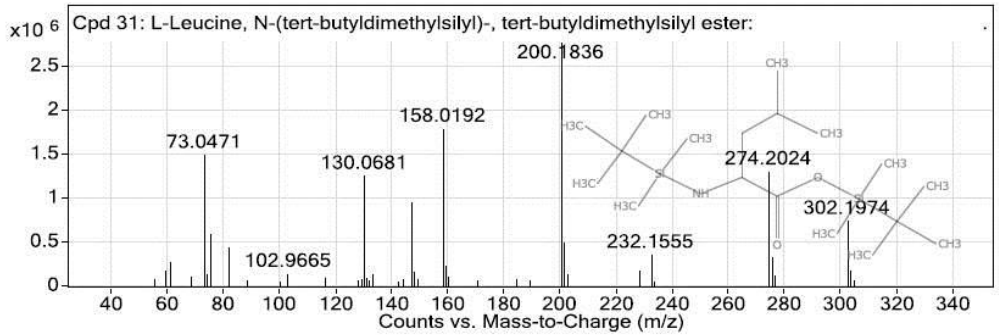


Leucine

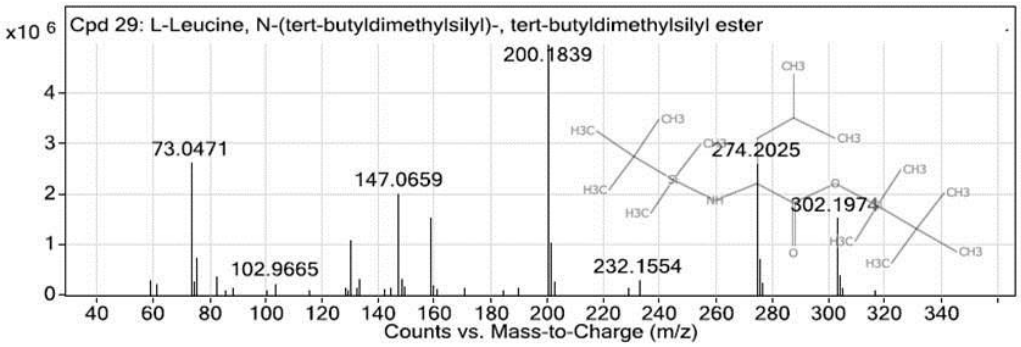
10 x difference in scale



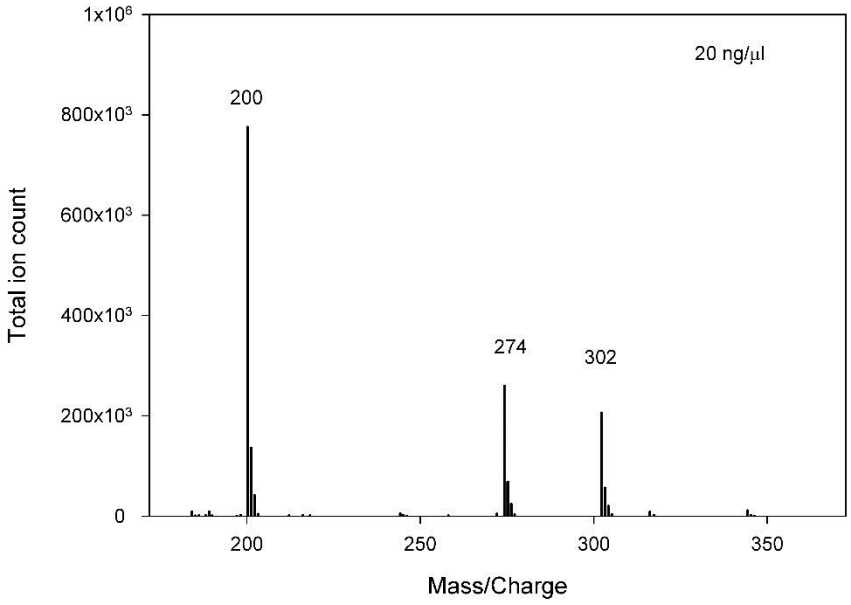
10 ng/µl standard in pyridine



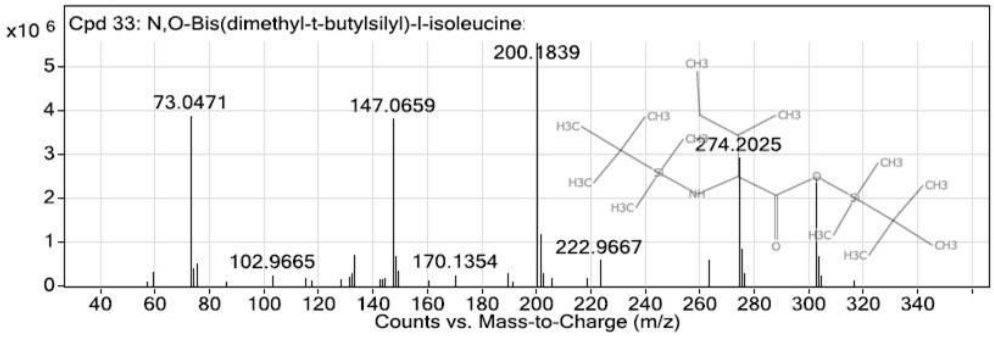
20 ng/µl standard in pyridine



Isoleucine

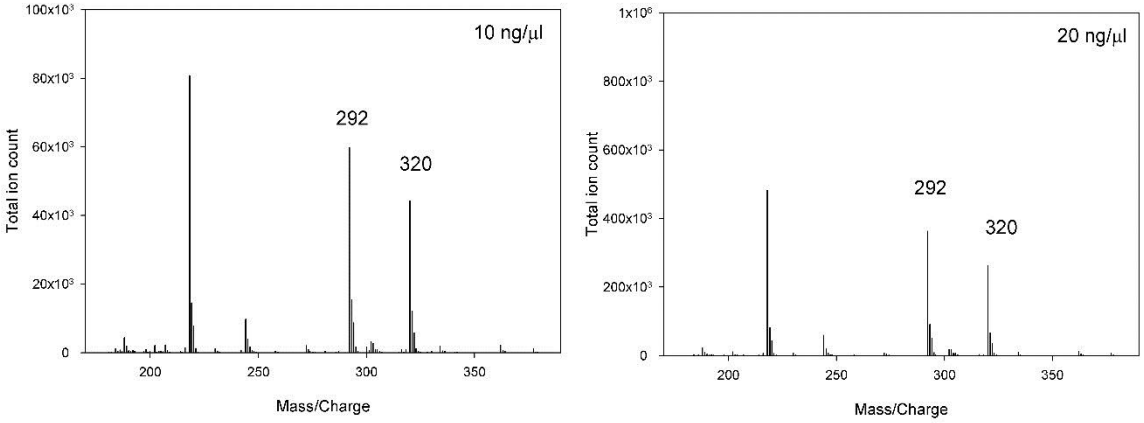


20 ng/µl standard in pyridine

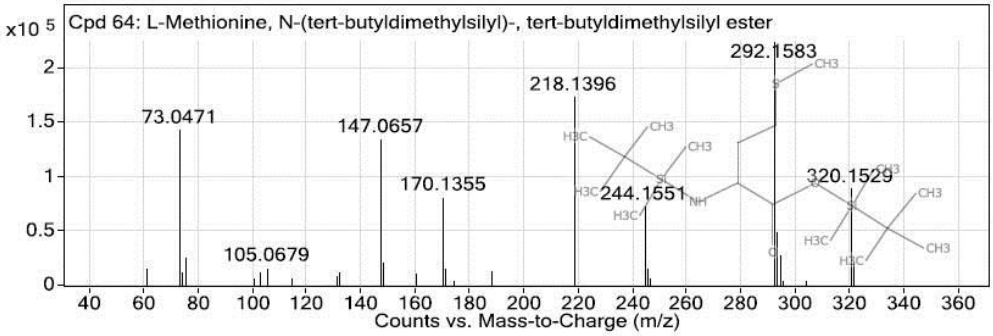


Methionine

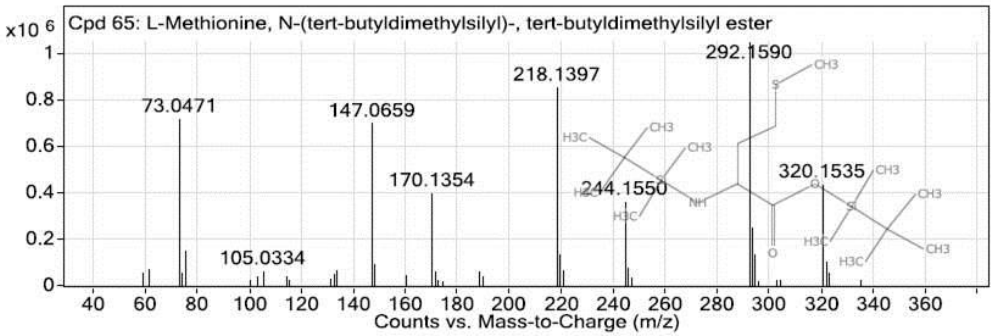
10 x scale difference



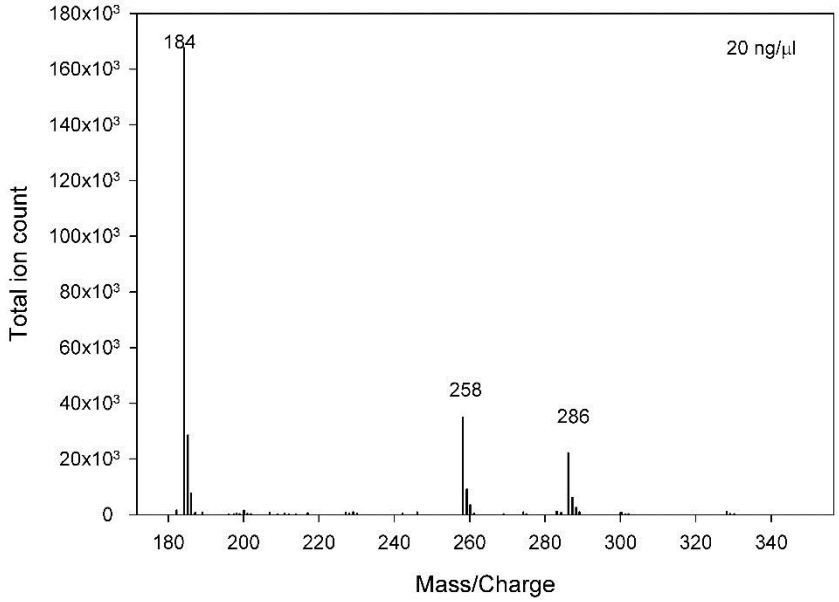
10 ng/µl standard in pyridine



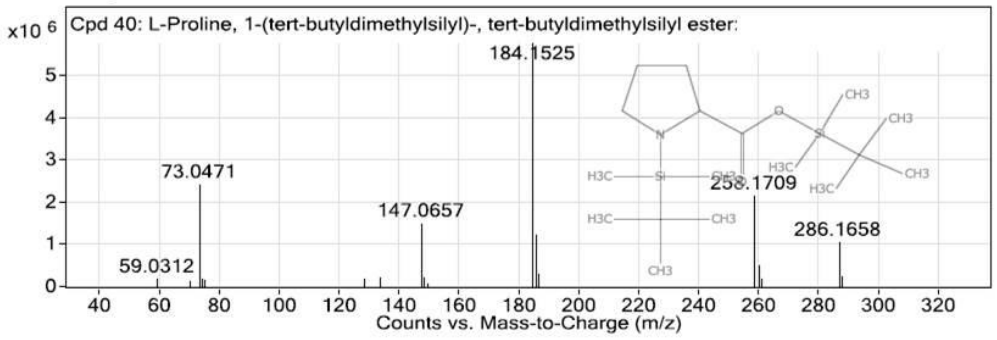
20 ng/µl standard in pyridine



Proline

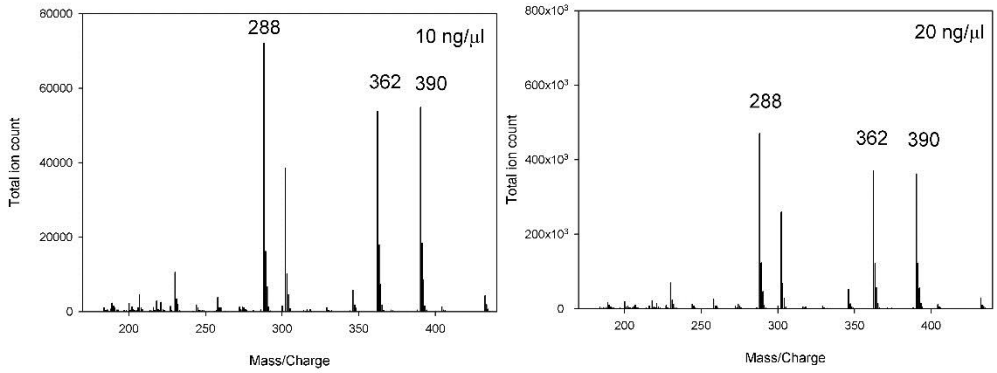


20 ng/µl standard in pyridine

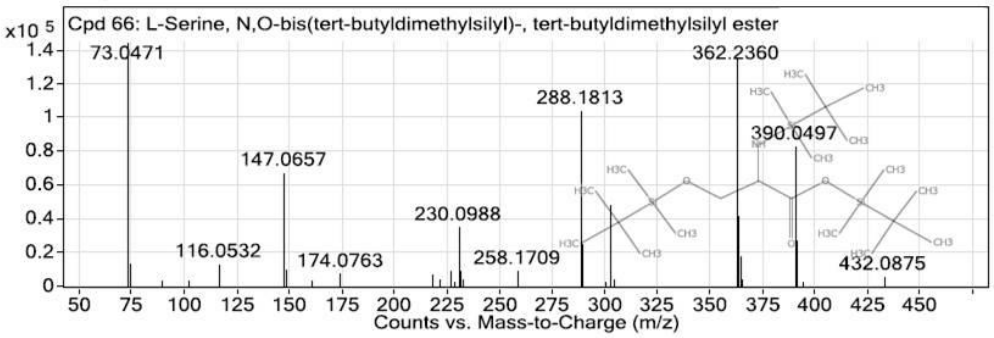


Serine

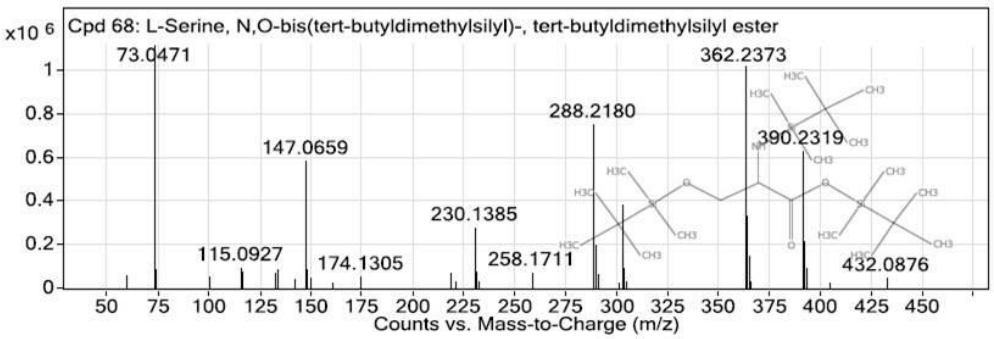
10 x scale difference



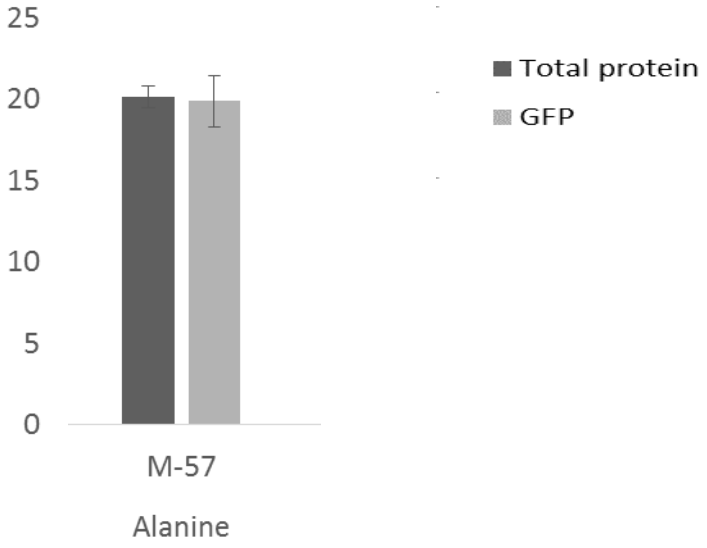
10 ng/µl standard in pyridine

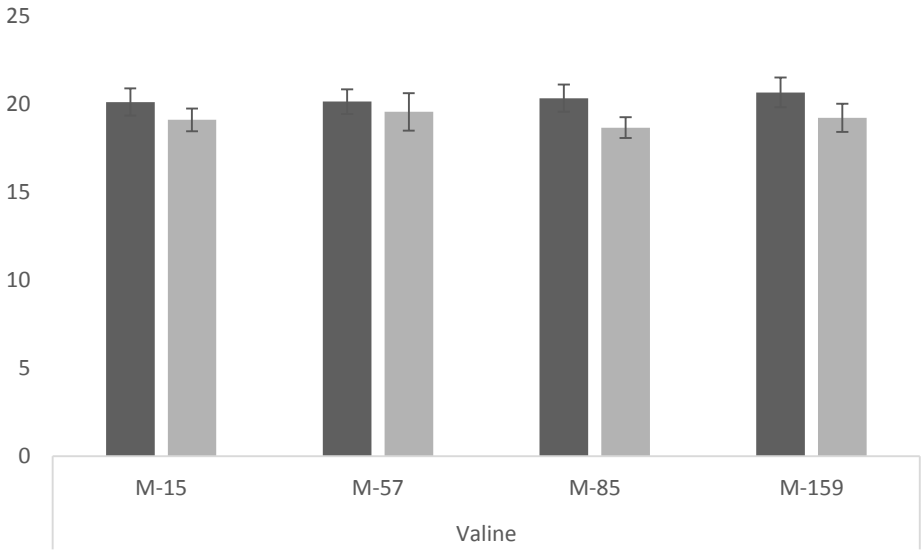


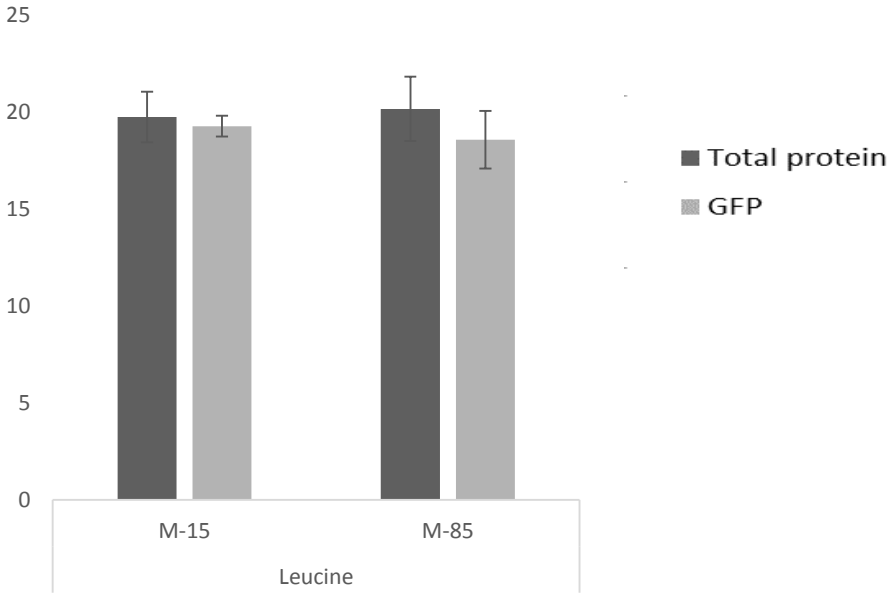
20 ng/µl standard in pyridine

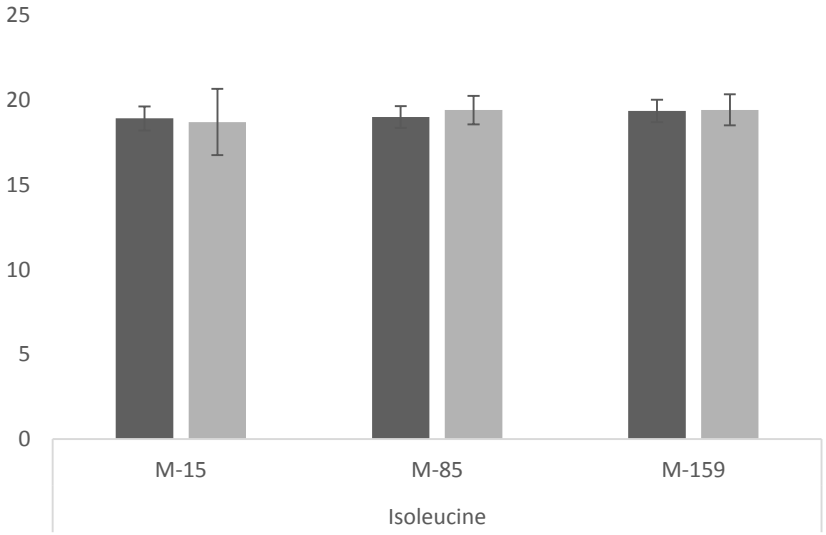


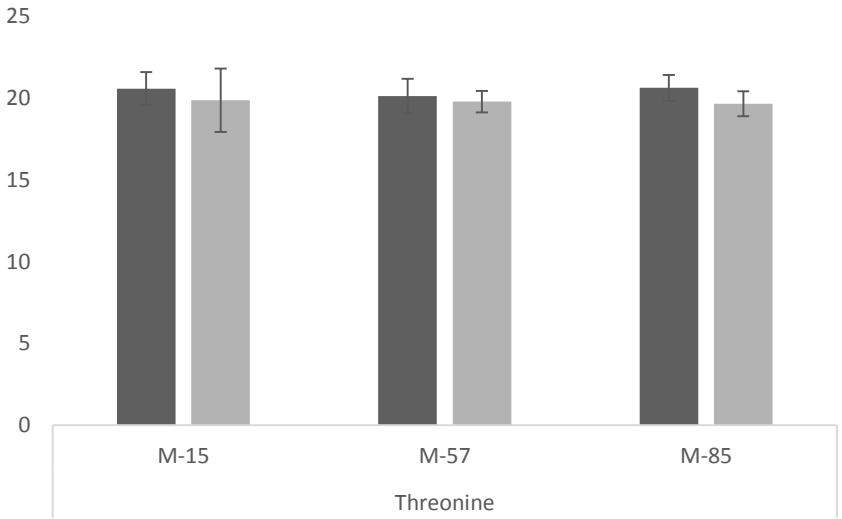
Appendix 16: **Comparison of average ^{13}C percentage derived from amino acid fragments detected from purified GFP samples and the same fragments from the total protein fraction.** Data are presented as the mean ± SD from at least three and a maximum of 15 biological replicates.

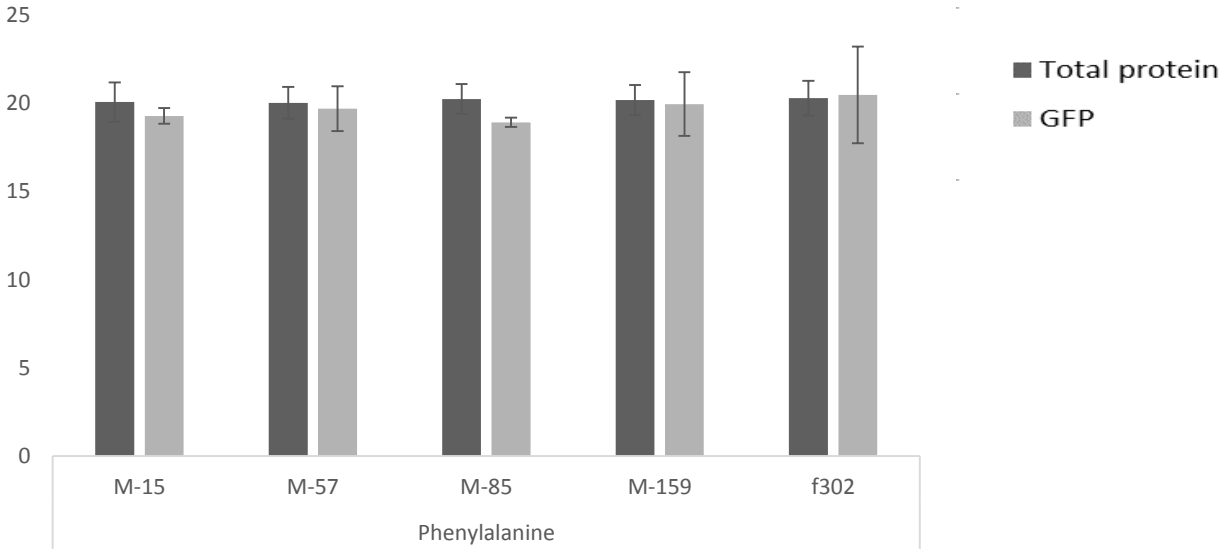


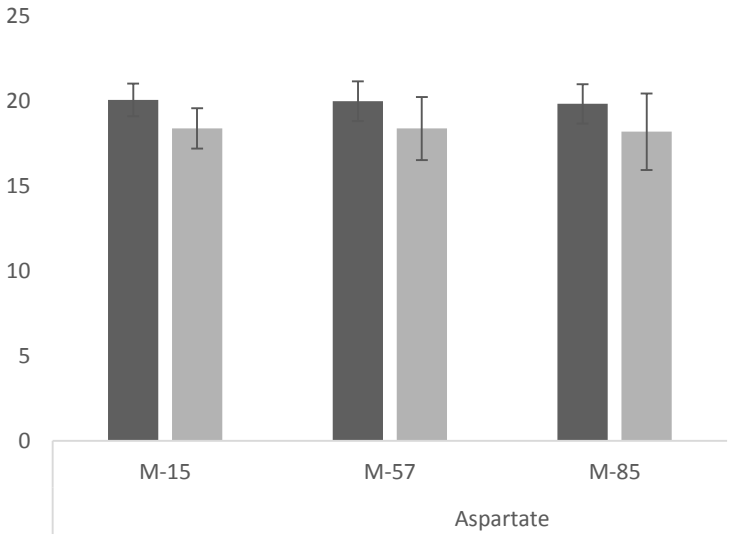


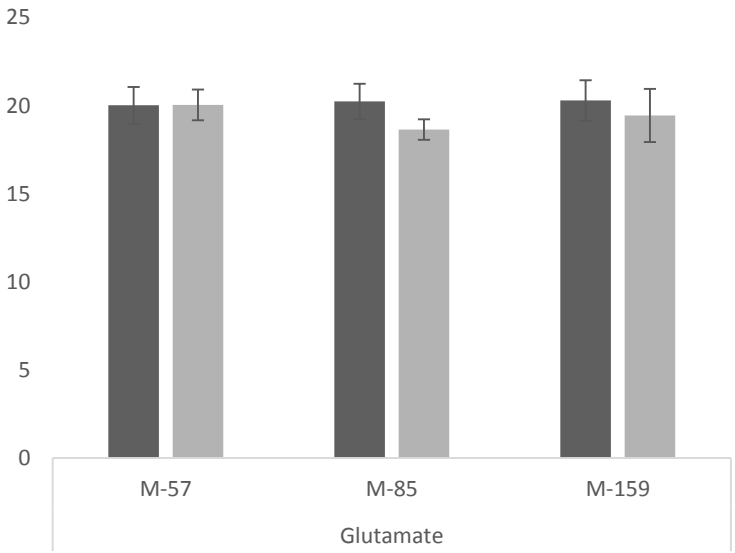

25
20
15
10
5
0
Total protein
GFP
M-15
M-85
Leucine

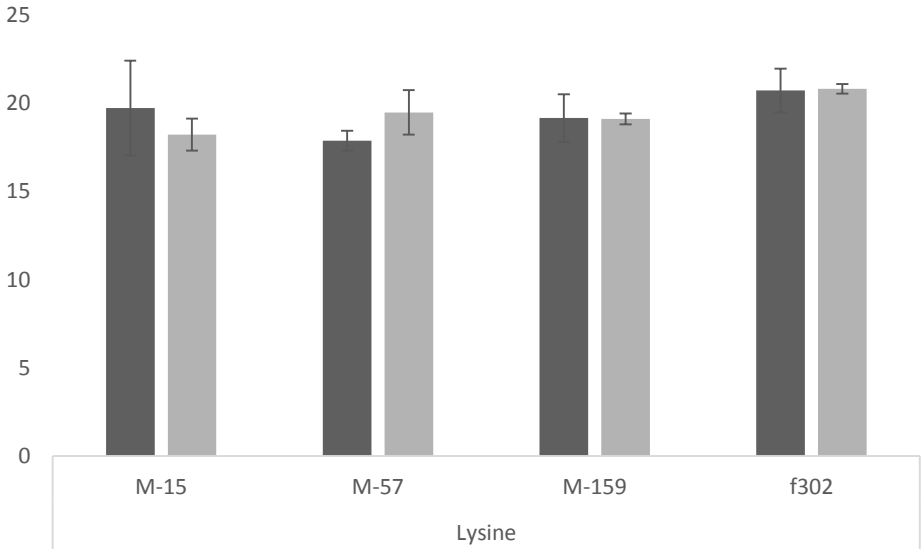

25
20
15
10
5
0
M-15
M-85
M-159
Isoleucine

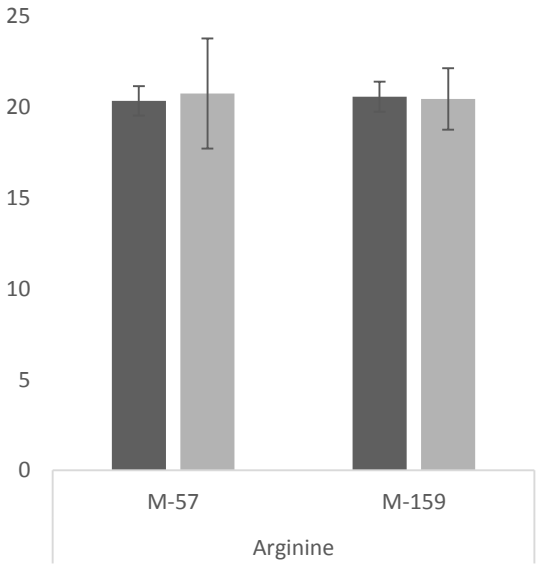

25
20
15
10
5
0
M-15
M-57
M-85
Threonine

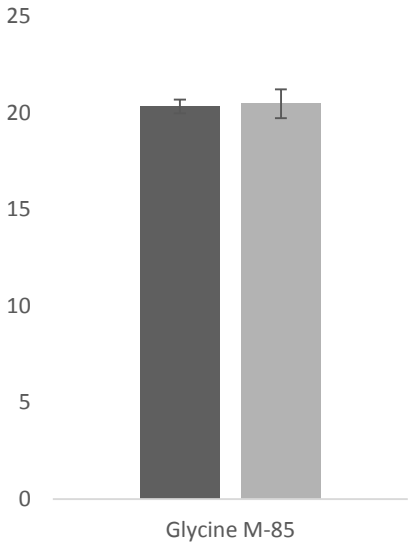

25
20
15
10
5
0
M-15
M-57
M-85
M-159
f302
Phenylalanine
Total protein
GFP

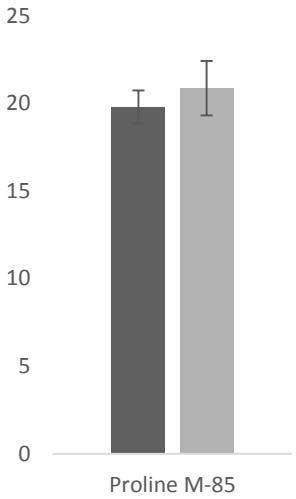

25
20
15
10
5
0
M-15
M-57
M-85
Aspartate


25
20
15
10
5
0
M-57
M-85
M-159
Glutamate


25
20
15
10
5
0
M-15
M-57
M-159
f302
Lysine

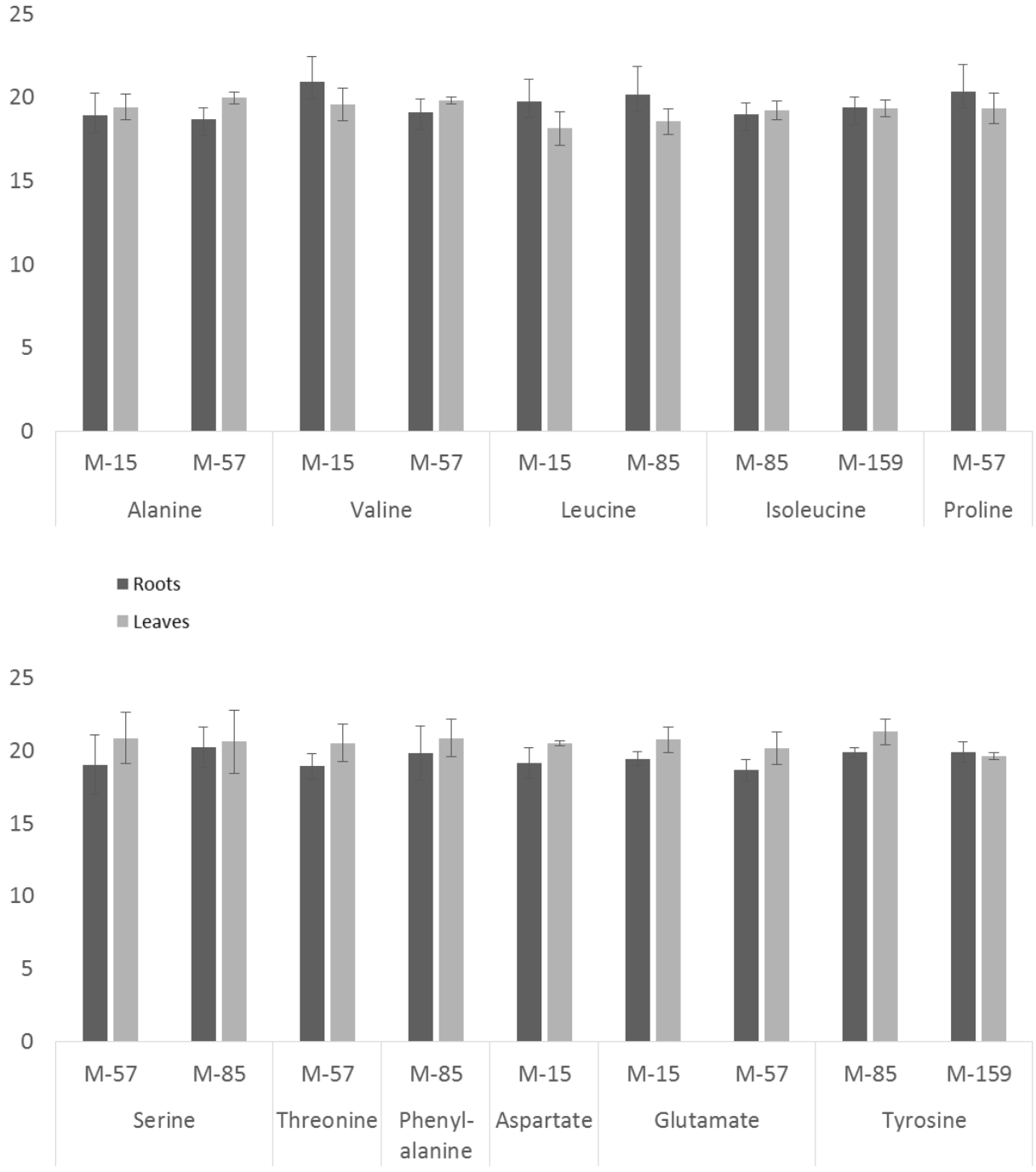

25
20
15
10
5
0
M-57
M-159
Arginine


25
20
15
10
5
0
Glycine M-85

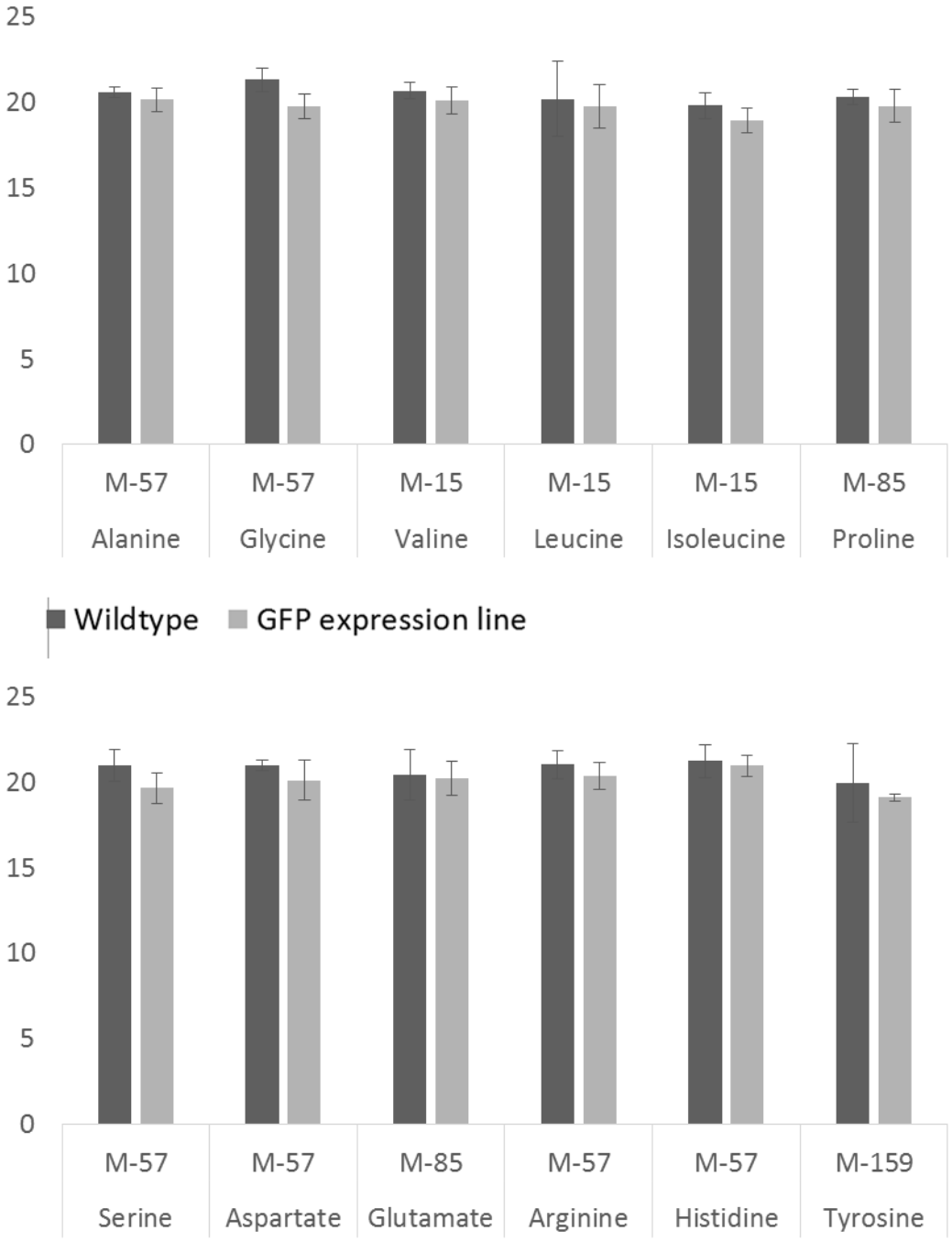

25
20
15
10
5
0
Proline M-85


Total protein
GFP

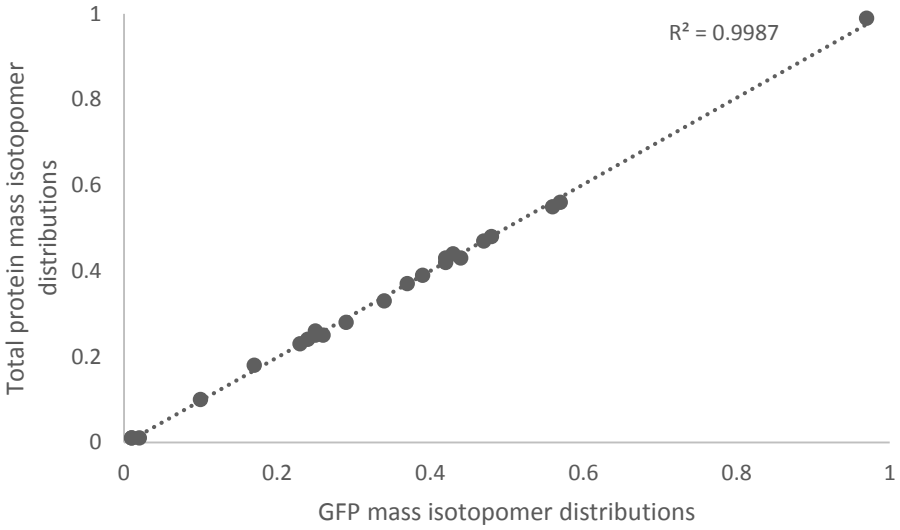
Appendix 17: **Comparison of fractional abundances of ^{13}C in amino acid fragments detected from the total protein fractions from root and shoot tissue in Arabidopsis seedlings.** Data are presented as the mean ± SD from at least three and a maximum of 15 biological replicates.



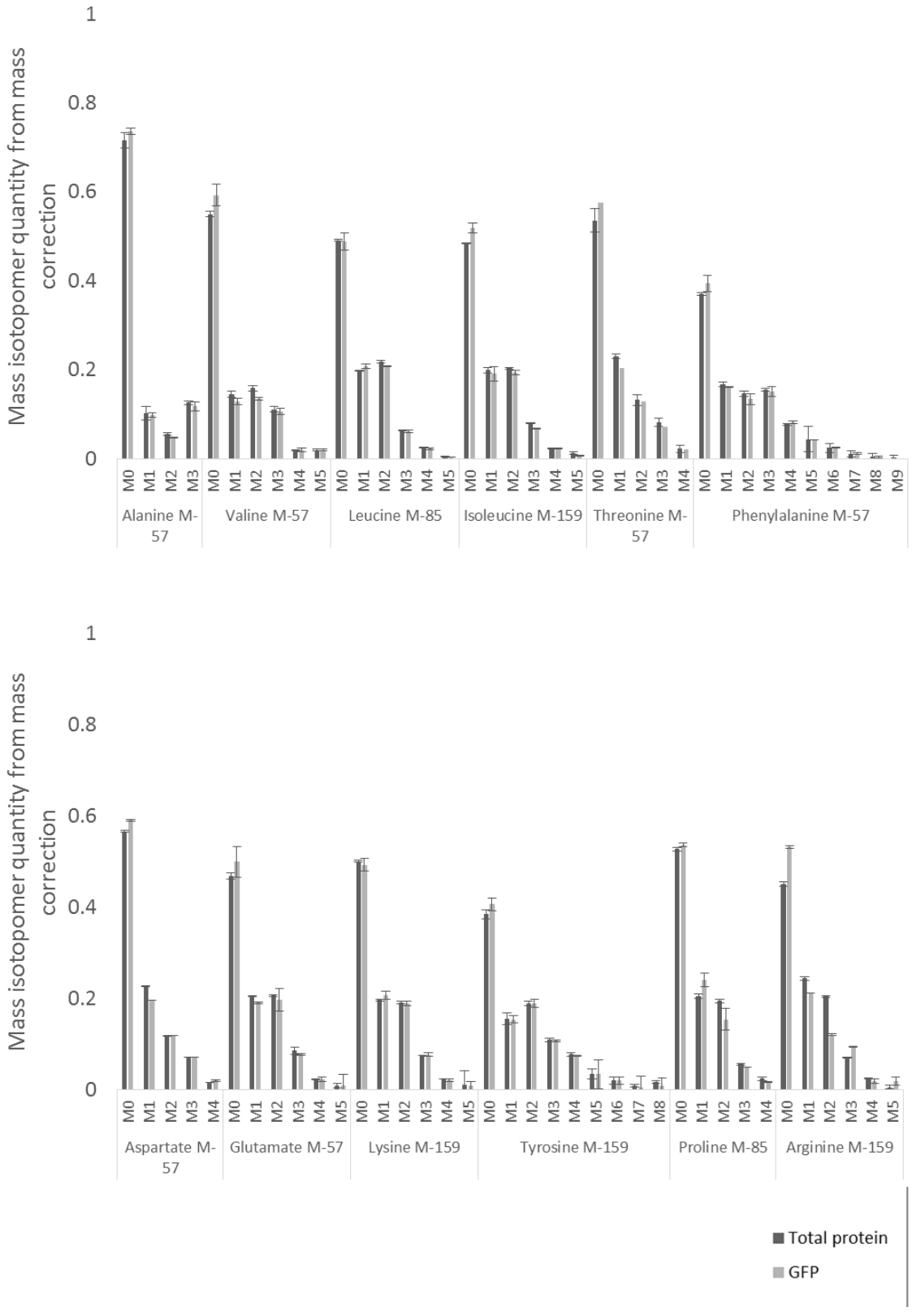
Appendix 18: **Comparison of fractional abundance of ^{13}C detected from total protein fractions from wildtype seedlings and seedlings with constitutive GFP expression.** Data are presented as the mean ± SD from at least three and a maximum of 15 biological replicates.



Appendix 19: **Comparison of the mass isotopomer distribution of amino acids derived from GFP and total protein in E. coli. Reanalysis of data presented in Shaikh et al. (2008)**



Appendix 20**: Mass isotopomer distributions in amino acid fragments derived from hydrolysed total protein fraction and GFP from Arabidopsis seedlings with standard deviations.**



Appendix 21: **Statistical comparison of average ^{13}C enrichment in protein derived from isolated root hairs and root total protein fraction** (Benjamini and Hochberg, 1995)**.** The Benjamini-Hochberg procedure was used to define a false discovery rate for the set of data, with 0.05 as the accepted false discovery rate and the results that were significant in this analysis are shown in Table A. The analysis (Table B) led to identification of three cases where the processing of the data suggested the variation may not be significant despite the Student's t-test producing a value below the cut off. The calculation of false discovery rates was done by using the total number of tests (27), from which selected final results are shown in Table A.

Table A: **Results from analysis of ^{13}C enrichment in amino acid fragments from isolated root hairs and total protein fraction adjusted for the false discovery rate.** P-values which are below the cut off value 0.05 have been highlighted. Benjamini-Hochberg procedure yielded a false discovery rate of 0.0167 and p-values below this (significant variation) are highlighted with black borders. Values which additionally are smaller than a cut off obtained from a more conservative Bonferroni correction (calculated by dividing the rate 0.05 by the number of tests) are highlighted in bold.

Amino acid		P-value	Adjusted p-value
Valine	M-15	0.043561	0.098012
Methionine	M-159	0.012704	0.038113
	M-85	**0.00059**	0.004747
Serine	M-57	0.030921	0.075896
Phenylalanine	M-85	0.026149	0.070602
	M-159	0.003688	0.014226
	M-15	**0.000613**	0.004747
Aspartate	M-57	**0.000551**	0.004747
	M-85	**0.000773**	0.004747
	M-15	**0.007499**	0.025308
Glutamate	M-57	**0.000879**	0.004747
	M-85	0.003246	0.014226

Table B: **The Benjamini-Hochberg procedure used to control the false discovery rate.**

Conventional metabolic flux analysis experiments contain steps for statistical assessment of the reliability of the data produced through a labelling experiment and the resulting models, most importantly by fitting the data by Monte Carlo simulations and estimating confidence intervals (Antoniewicz et al., 2006; Masakapalli et al., 2013). Here the information on fractional abundances compared between different sample types has not been subjected to such processes and other statistical approaches were included in the processing of the data in order to confirm which variation was statistically significant between the sets of biological data.

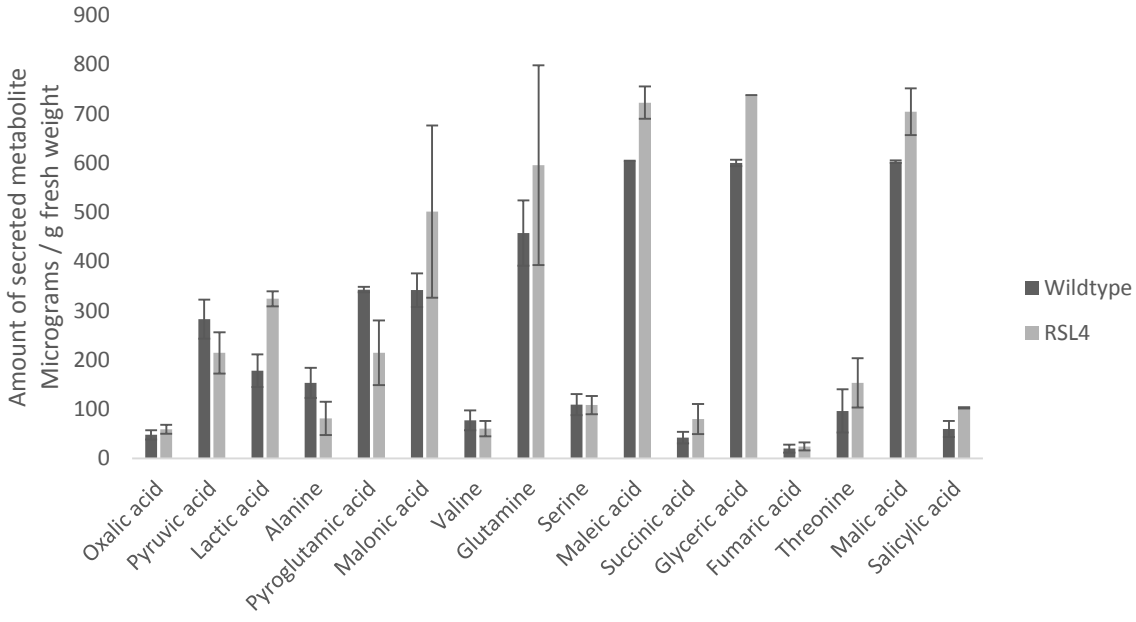
In order to assess the statistical significance of variation detected between data points in all of the datasets, a Student's t-test was used to identify any variation that would have a p-value below a 0.05 cut off value. The simple statistical analysis used a paired two sample t-test for means assuming unequal variances between two types of data. For a dataset consisting of data on fractional abundances of ^{13}C in amino acid fragments derived from a protein fraction in wildtype plants and those with RSL4 expression, the data subjected to the analysis consisted of the fractional abundances calculated for a specific amino acid fragment from each biological replicate from the two sets of data.

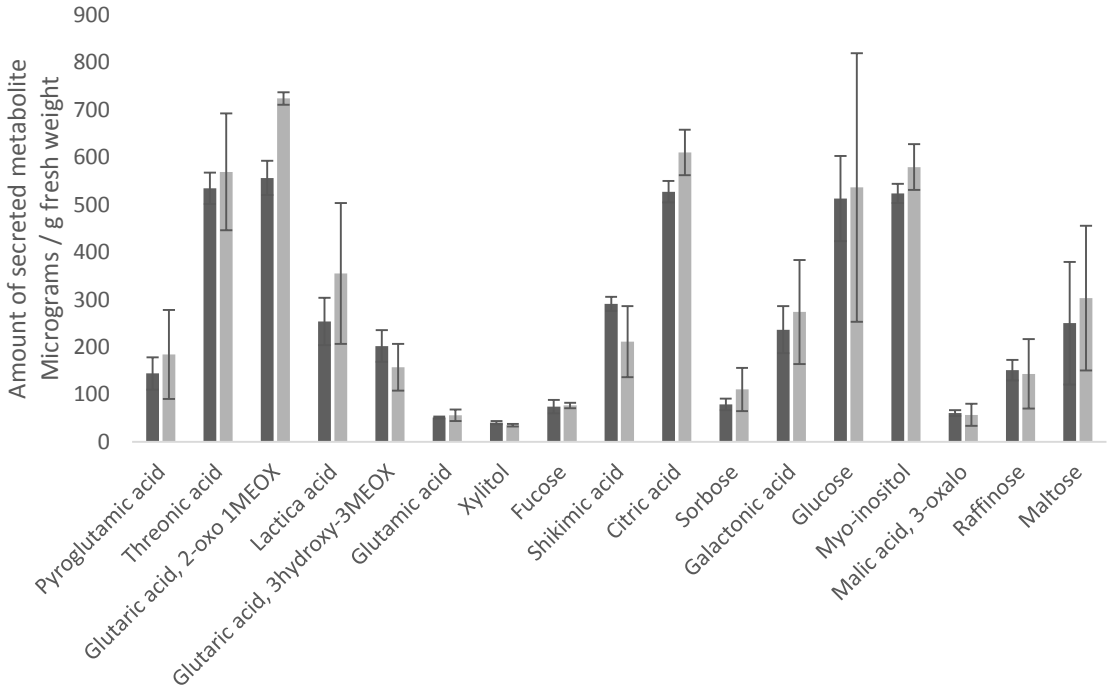
In the few instances where the results are referred to as significant variation in Chapter 6, the Benjamini and Hochberg procedure has been applied on the p-values to control false discovery rate. Vast majority of the variation with p-values lower than 0.05 were significant according to both the Benjamini and Hochberg procedure and also a more conservative Bonferroni correction. For the variation that was identified as potential false positives, other fragments for the same amino acid showed significant variation in all cases except for one out of the three. Since the experiments used uniformly labelled substrates, results suggest that the false discovery rate may have been set high for an initial experiment and confirming the significance of these results should be considered in future work (see 6.2.2, Figure 35).

The p-values obtained for each set of amino acid fragment data pairs were sorted in order from the smallest to the largest and i-values were assigned based on their rank. The values were compared to the factor (i/m)Q, where m is the total number of tests and Q is the chosen acceptable false discovery rate, which can be varied depending on the requirements of the experiment and effect of false positives. The procedure aims to find the largest p-value which satisfies the criteria P<(i/m)Q. All the p-values that are smaller than this are considered significant, regardless of the p-value (bold). An alternative is to calculate adjusted p-values (Table A), but they were not used for assessing statistical significance of these results. Determining a cut off value based on the Benjamini-Hochberg principle yields a value of 0.0167 and leads to identical results.

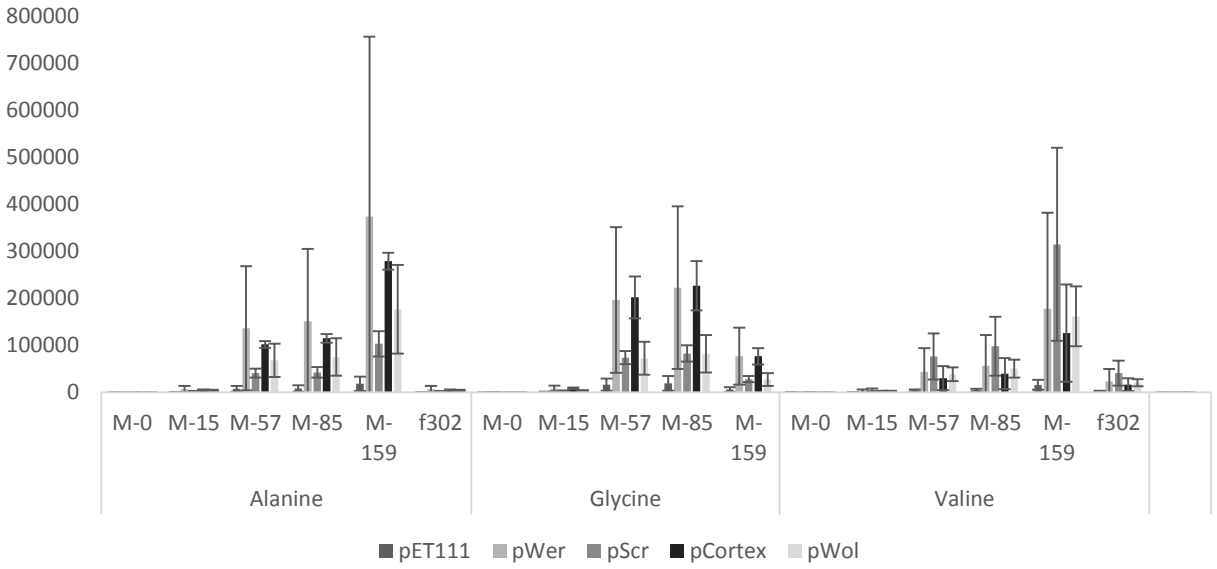
m	Q	P-value	i	(i/m)Q	P<(i/m)Q?
27	0.05	**0.000551**	**1**	**0.001852**	**yes**
		0.00059	**2**	**0.003704**	**yes**
		0.000613	**3**	**0.005556**	**yes**
		0.000773	**4**	**0.007407**	**yes**
		0.000879	**5**	**0.009259**	**yes**
		0.003246	**6**	**0.011111**	**yes**
		0.003688	**7**	**0.012963**	**yes**
		0.007499	**8**	**0.014815**	**yes**
		0.012704	**9**	**0.016667**	**yes**
		0.026149	10	0.018519	no
		0.030921	11	0.02037	no
		0.043561	12	0.022222	no
		0.060241	13	0.024074	no
		0.069272	14	0.025926	no
		0.078287	15	0.027778	no
		0.084635	16	0.02963	no
		0.093488	17	0.031481	no
		0.102532	18	0.033333	no
		0.109318	19	0.035185	no
		0.142361	20	0.037037	no
		0.197163	21	0.038889	no
		0.246575	22	0.040741	no
		0.24674	23	0.042593	no
		0.298756	24	0.044444	no
		0.341528	25	0.046296	no
		0.371406	26	0.048148	no
		0.476212	27	0.05	no

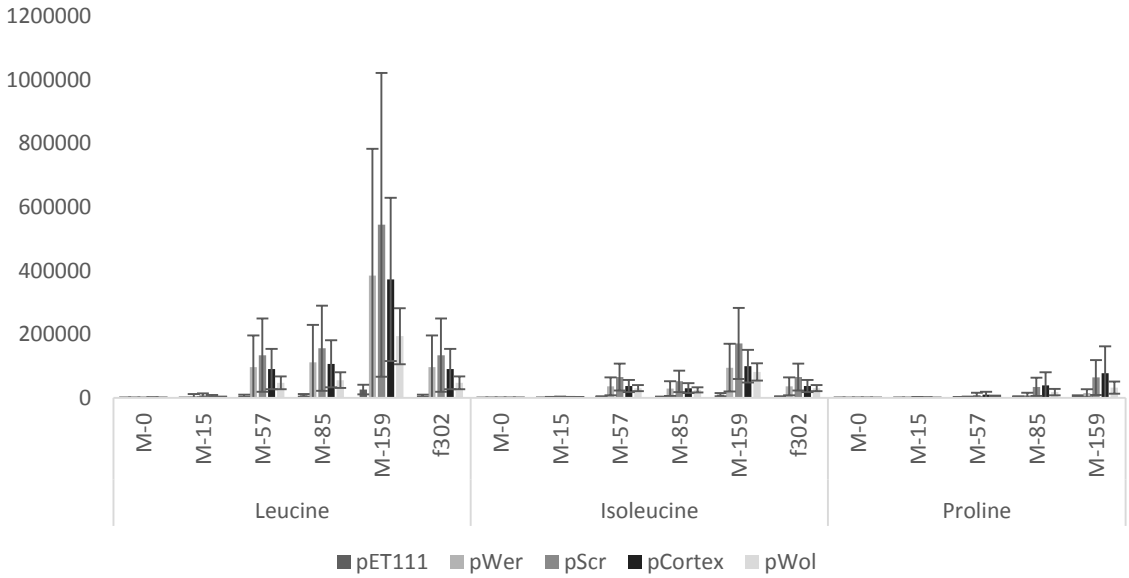
Appendix 22: **Comparison of levels of metabolites detected from exudates from the wildtype Columbia seedlings and seedlings with RSL4 over expression.** Data are displayed as the mean and standard deviation of amounts of secreted metabolite measured from the exudate. A minimum of four samples from biological replicates were used for each data point. Student's t-test corrected for false discovery rate by Benjamini-Hochberg procedure was used to assess statistical significance of differences between the datasets.

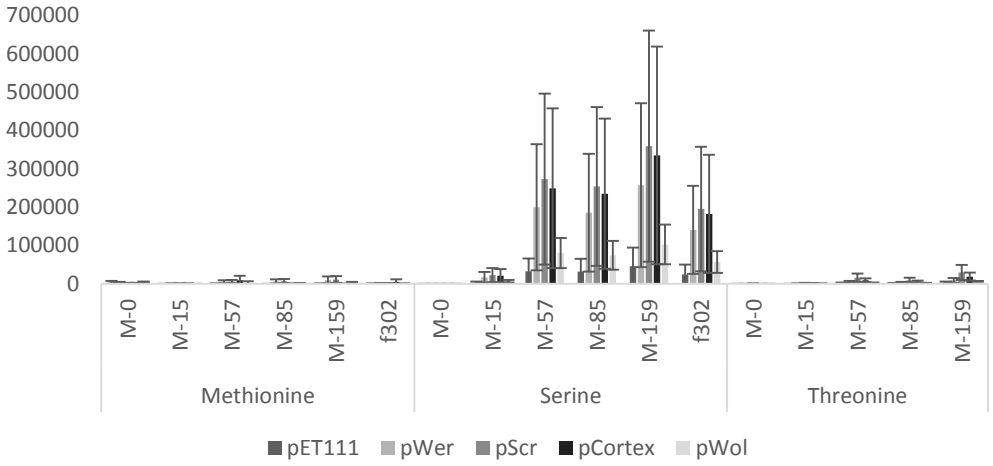


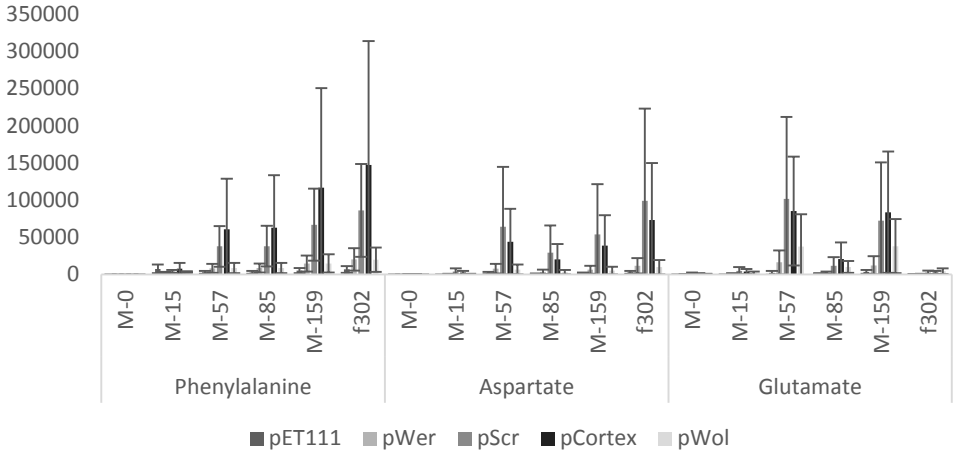


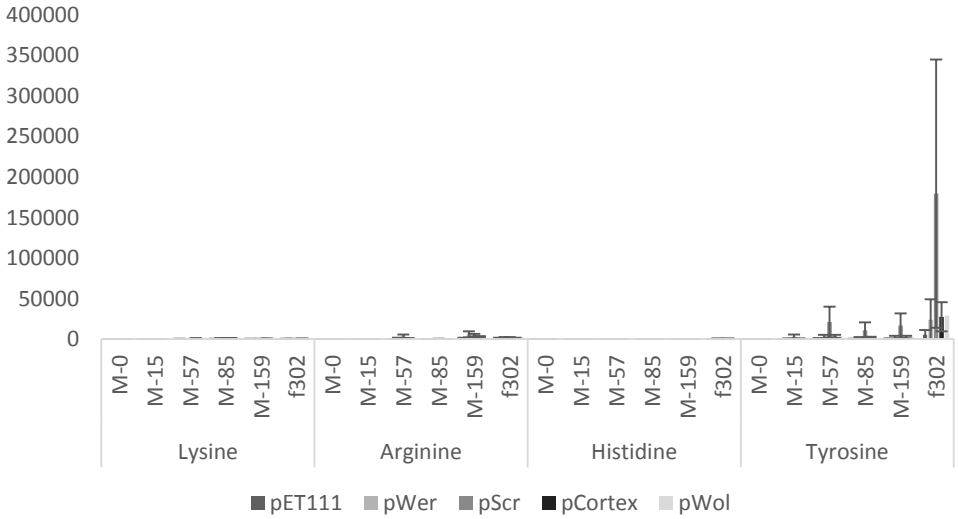
Appendix 23: Levels of amino acid fragments detected from GFP purified from specific cell types of Arabidopsis seedlings from 5 cultures of 30-50 seedlings combined in each sample, mass peak corresponding to fragment containing no ^{13}C substitutions. Each value is the mean ± SD from a minimum of four independent samples.






700000
600000
500000
400000
300000
200000
100000
0
M-0
M-15
M-57
M-85
M-159
f302
Methionine
Serine
Threonine
pET111
pWer
pScr
pCortex
pWol


350000
300000
250000
200000
150000
100000
50000
0
M-0
M-15
M-57
M-85
M-159
f302
Phenylalanine
Aspartate
Glutamate
pET111
pWer
pScr
pCortex
pWol


400000
350000
300000
250000
200000
150000
100000
50000
0
M-0
M-15
M-57
M-85
M-159
f302
Lysine
Arginine
Histidine
Tyrosine
pET111
pWer
pScr
pCortex
pWol