COPYRIGHT AND CITATION CONSIDERATIONS OF A THESIS/ DISSERTATION

You should include the following information in your bibliography, the exact style will vary according to the citation system you are using:

Name of author

Year of publication, in brackets

Title of thesis, in italics

Type of degree (e.g. D. Phil.; Ph.D. or M.Sc.)

Name of the University

Country

Website

Date, accessed

Example

Ndlovu, T. (2012) *Electrochemical detection of organic and inorganic water pollutants using recompressed exfoliated graphite electrodes*. D. Phil. University of Johannesburg: South Africa. Retrieved from: https://ujdigispace.uj.ac.za (17/7/2013).

Metabolomic studies of biotransformation-related changes in plant metabolism in response to isonitrosoacetophenone treatment.

by

Ntakadzeni Edwin Madala

Thesis

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

in

BIOCHEMISTRY

in the

FACULTY OF SCIENCE

at the

UNIVERSITY OF JOHANNESBURG

SUPERVISOR: PROF IA DUBERY

CO-SUPERVISORS: DR LA PIATER and PROF PA STEENKAMP

NOVEMBER 2012

To my Mother:

Ntsumbedzeni Selinah Makumbane - Madala

“Ndo livhuwa nga maanda, ngavhe Mudzimu a tshi vha tonda nga manwe maduvha a vhutshilo!”

Ndaa!

Acknowledgements

I would like to thank the following people for their individual contributions in my social and academic life:

- My supervisor, Prof. Ian Dubery is thanked for allowing me an opportunity of a lifetime, to study and pursue my lifelong dream of having a Ph.D. one day. Your guidance/supervision, from my early B.Sc.Hons, M.Sc. and now Ph.D., is really appreciated. I thank you for sharing your knowledge with me and by walking and working with me when I needed supervision the most.
- My co-supervisor, Dr. Lizelle Piater, I get very emotional thinking of all the effort you made in my study, together we learnt metabolomics, and never in a single day you failed to impress me with your biochemistry knowledge. Your strong command of the English language is the reason this thesis is the way it is. I can write another thesis describing your contributions. Thank you Lizelle! I wholeheartedly appreciate all your support and guidance.
- My second co-supervisor, Prof. Paul Steenkamp, the first day you became part of this study, I knew I was about to deliver a killer study. Your knowledge of analytical chemistry is amazing. Your technical assistance with the UHPLC-MS analyses is what made this thesis to be a reality now. Paul, I thank you for all the wonderful, educative and insightful discussion we had.
- Mrs Marina Dubery, I thank you for all the cells that you prepared for us. All the conclusions of this thesis wouldn't be possible if I didn't have healthier cells.
- My laboratory friends and colleagues, especially those working on the field of metabolomics, Fidele Tugizimana, Tarryn Finnegan, Leandri Rassouw and Heino Heyman. I truly thank you for all the advises and discussions you shared with me throughout this journey. Whenever things were not working, you were the first to come for rescue. Fidele I personally thank you for everything, courage! Lindy, Londi, Maureen, Peggy, Furaha, Makhadzi, Mulisa, Mashavhela, Paseka, Sandile (SBZ), Mosotho, Swiki, and Venessa O'Neil, I truly thank you. My personal friends, Khumbudzo, Tuwi, Litshani, Ntiyi, Fulu, Murendeni, Dzungi, Gwadamani, Mashudu, Pfarelo, Shoni, Taki, Lufuno and Mazwi thank you all for your friendship and for walking with me through this journey.
- Boitumelo, words can't express fully my sincere gratitude, I wholeheartedly thank you for all your love and support.
- To my family, my siblings, Livhuwani, Phumudzo and Mukundisi, thank you for your individual love. To my mother, whom the thesis is dedicated to, thank you for your love and support, especially allowing me to go to school whilst other parents were forcing their children to work.
- University of Johannesburg and the NRF are thanked for their lucrative scholarships I received throughout my studies.
- To God Almighty, I am constantly encouraged by your promise in Joshua 1: 9, "Have I not commanded you? Be strong and courageous. Do not be afraid; do not be discouraged, for the LORD your God will be with you wherever you go".

TABLE OF CONTENTS

Preface

This thesis concerns a study of the effect of isonitrosoacetophenone on plant metabolism. Three different systems were investigated; cultured tobacco and sorghum cells as well as *Arabidopsis thaliana* plants, and a metabolomic approach was followed. Unlike most scientific studies, metabolomics is a discipline which is not driven by a specific hypothesis, but rather by the obtained data to add scientific insights to the topic under investigation. As such, the current study lacks a definite overarching hypothesis, but specific objectives were outlined and answered in each experimental chapter. This thesis is therefore presented as a compilation of nine chapters in which experimental/research work is described in Chapter 3-8. It is important to note that each chapter is presented in accordance with the guidelines for the respective journal in which the corresponding manuscript was published or submitted to. The outline of the thesis is as follows:

Chapter 1: Summary and aims of the study.

Chapter 2: Literature review.

Chapter 3: The short and long of it: Shorter chromatographic analysis suffice for sample classification during UHPLC-MS-based metabolic fingerprinting. **NE Madala**, F Tugizimana, PA Steenkamp, LA Piater, IA Dubery, *Chromatographia* (2013) accepted: DOI: 10.1007/s10337-012-2336-z. (Special issue 'Chemometrics in Chromatography').

Chapter 4: Collision energy alteration during mass spectrometric acquisition is essential to ensure unbiased metabolomic analysis. **NE Madala**, PA Steenkamp, LA Piater, IA Dubery, *Analytical and Bio-analytical Chemistry* (2012) 404: 367-372.

Chapter 5: Isonitrosoacetophenone induces perturbations in the metabolic status of tobacco cells. **NE Madala**, PA Steenkamp, LA Piater, IA Dubery (2012) To be submitted to *Phytochemistry.*

Chapter 6: Metabolic changes induced by isonitrosoacetophenone in *Arabidopsis thaliana* plants results in an enhanced defensive environment. **NE Madala**, PA Steenkamp, LA Piater, IA Dubery (2012) To be submitted to *Journal of Plant Physiology.*

Chapter 7: Deciphering time-dependant trends in metabolomic data from elicited plant cells using multivariate statistical models. **NE Madala**, S Halouska, PA Steenkamp, LA Piater, IA Dubery (2012) Submitted to *Analytical Biochemsitry*.

Chapter 8: Bioconversion of isonitrosoacetophenone (2-keto-2-phenyl-acetaldoxime) in tobacco cell suspensions. **NE Madala**, PA Steenkamp, LA Piater, IA Dubery, *Biotechnology Letters* (2012) 34: 1351-1356.

Chapter 9: General conclusion.

Chapter 1: Summary

Plants are constantly faced with numerous biotic and abiotic stress factors and, unlike mammals, lack adaptive immunity and thus have to develop a way to overcome these devastating obstacles. One hallmark of plant mechanisms to defend themselves against these stress factors is based on a sophisticated innate immunity. Plant innate immunity, or basal resistance as it used to be referred to, can be divided into two facets, namely the inducible (active) and constitutive (passive) response. Both these forms of resistance can further be divided into structural and chemical defense responses. By nature, plants possess some structural and chemical components which allow them to fend off attacking stress factors such as pathogens, insects and herbivores. These components include, amongst others, cell wall layers such as cuticles and other chemicals which make the host plant unpalatable to a wide variety of herbivores and/or impenetrable to pathogens. However, during stress encounters, plants are also capable of inducing these chemicals and cell wall strengthening compounds, all of which are aimed at defense against the attacking stress factor. To a certain extent, the same strategy is also utilized to fight against abiotic stress factors such as drought and UV radiation.

The current study focussed mainly on the chemically-inducible defense responses. Here, metabolomics as a new emerging field of study was evaluated to assess the impact of a chemical inducer of defense, isonitrosoacetophenone (INAP), on the metabolome of different plants. INAP, or 2-keto-2-phenyl-acetaldoxime, is a structural analogue of 4-(3-methyl-2-butenoxy)-isonitrosoacetophenone, an oxime-containing stress metabolite/phytoalexin, (4-(3-methyl-2-butenoxy)-isonitrosoacetophenone or citaldoxime) which was previously shown to accumulate in citrus peel undergoing oxidative stress due to gamma radiation treatment as well as resulting in an anti-fungal environment. Oxime functional groups are rare in natural products and the existence thereof is restrictively found in plants which are capable of producing specialised molecules known as glucosinolates and cyanogenic glycosides. Hitherto, the accumulation of citaldoxime and other oxime molecules in non-cyanogenic plants is not fully understood. It was from this notion that the current study stemmed. With the aid of UHPLC-MS and multivariate statistical data models, a comprehensive study to partially understand the accumulation of oximes in non-cyanogenic plants in comparison to cyanogenic counterparts was conducted and shown to successfully elucidate the metabolic fate and metabolic- / defense- inducing ability of INAP.

Before the full-scale metabolomics study, method development procedures to better suite the outcomes of the current study were carried out. Here, optimizations of the different aspects of UHPLC-MS were required. Firstly, the chromatographic method was optimized by changing different factors such as column type, solvent systems and chromatographic run times. With the aid of chemometric data models (Principle component analysis (PCA) and orthogonal projections to latent structure discrimate analysis (OPLS-DA)), the effect of chromatographic separation length was evaluated and the outcomes of this study are reported in **Chapter 3**. Contrary to public perception, it was shown that the length of UHPLC-MS chromatography does not adversely affect the outcome of the analysis. It was shown that the statistical models managed to separate the samples from different biological groups (controls and INAP-treated) in a similar manner regardless of the length of chromatography. The outcomes of this study are expected to have a positive impact on all scientists working in the field of metabolomics by eliminating the concern of loosing information when shorter chromatographic separation is employed. However, this study does not recommend the use of unqualified chromatographic length or conditions in general. The application of shorter chromatography will be feasible, for example, in the field of chemotaxonomy where the main intention is to classify the samples based on the chemical constituents. However, in cases where the objective is to fully understand the contributors to the differences (i.e. chemicals) between samples of varying biological background, optimization in chromatographic length is encouraged. It must also be kept in mind that the findings of this chapter could somehow be attributed to the type of instrument used (UHPLC) and its mass component with the ability to measure the monoisotopic mass of molecules with high accuracy (≤ 5ppm).

Optimization of the MS components was also investigated and here, changes in the collision energy required for molecule fragmentation during MS acquisition was evaluated. The outcomes of this study, as presented in **Chapter 4**, shows that changes in collision energy is essential to ensure comprehensive coverage of the metabolome. Here, the same statistical models as in Chapter 3 were applied on data generated using different levels of collision energy, and it was found that such changes result in different sample grouping patterns. This chapter further shows that the identification of metabolites is somehow hindered by the level of collision energy utilized as the molecules become unstable at higher energy levels. It was thus concluded that the use of moderate energy levels is sufficient for MS-based metabolomics-related studies and optimization prior to full-scale analyses should be carried out. From this study it was further suggested that scientists working in the field of

metabolomics should take note of, and report on, the collision energy used when depositing findings into public databases so as to ensure that data is reproducible. This, in turn will ensure unbiased reporting. Both **Chapters 3 & 4** were designed in order to honour the Metabolomic Standards Initiatives which aim to encourage proper and accurate reporting during metabolomics studies.

Following the outcomes of the work presented in the latter two chapters, the effect of INAP on the metabolome of different plant systems (cells and tissue) was evaluated. Here, *Arabidopsis thaliana* (intact plants), *Nicotiana tabacum* (tobacco cell suspensions) and *Sorghum bicolor* (sorghum cell suspensions) were investigated. The reason for using these plants systems was mainly encouraged by the fact that they possess different genetic backgrounds, different secondary plant metabolic pathways, and thus different abilities of oxime metabolism. For instance, *Arabidopsis* was used because of its ability to metabolize oximes to form glucosinolates, whilst sorghum is able to produce cyanogenic glycosides (dhurrin to be specific) from oximes and lastly, tobacco represents a non-cyanogenic plant. Treating tobacco cell suspensions and *Arabidopsis* plants resulted in altered metabolic characteristics as predicted by PCA and OPLS-DA. The outcomes of these experiments are shown in two consecutive **Chapters 5 & 6**. From these chapters it was shown that INAP induced metabolites with known activity in stress responses. To be precise, metabolites with antimicrobial and antioxidant activity were shown to accumulate in INAP-treated samples. The total phenolics concentration was shown to increase in tobacco cells as a result of INAP treatment and, with the aid of diphenylpicryl-hydrazyl (DPPH)-TLC assay, was further shown to be capable of inducing an antioxidant environment *in vivo* (**Chapter 5**). These findings can be collectively used to explain the initial accumulation of INAP-related molecule (citaldoxime) in tissue undergoing oxidative stress. The results in *Arabidopsis* (**Chapter 6**) further show that INAP is also capable of inducing an acquired resistance-like response against the representative bacterial pathogen, *Pseudomonas syringae* pv *maculicola.* It was also shown that INAP is bio-transformed to a product similar to those believed to form as side products during cyanogenic glycoside biosynthesis. Together, these findings suggest possible roles of INAP during plant stress responses in general. Another highlight of **Chapters 5 & 6** was the use of the PUTMEDID_LC-MS workflows for metabolite identification and these analyses proved promising in the field of plant-based metabolomics.

In order to evaluate the effect of INAP on cyanogenic and non-cyanogenic plants, sorghum and tobacco cells were respectively used. UHPLC-MS data from the extracts of control and INAP-treated cells (at different time intervals) were qualitatively compared to each other using PCA, OPLS-DA-based shared and unique structures (SUS) plots, hierarchal cluster analysis (HCA) and Metabolic trees. From this study (**Chapter 7**), it was shown that INAP, as a xenobiotic compound, is better metabolized in sorghum than in tobacco cells. This conclusion was reached by comparing the differential clustering patterns between samples of biological background (controls and INAP-treated samples in different time intervals) from the two plant systems. It was thus concluded that INAP, as an oxime molecule, is hence better metabolized in cyanogenic than in non-cyanogenic plants as the response in the former is well coordinated when compared to the later system. It was further shown that the use of different/alternating statistical models to evaluate metabolic responses allows for comprehensive biological interpretation underlying the exhibited response with cross-validated outcomes.

Apart from the metabolic-inducing ability of INAP, this molecule was also shown to undergo biotransformation events characterized by a series of chemical modifications in the core of the INAP structure. The last chapter of this thesis (**Chapter 8**) entails the full characterization of biotransformation events which INAP was thought to undergo. With the aid of tandem mass spectromentry (MS/MS) and chemical intelligence software (MassFragment^{TM}) it was shown that this oxime molecule is modified by endogenous enzymatic processes of tobacco cells to result in a molecule with structural features similar to those of some endogenous anti-stress metabolites found in different plants. As such, INAP is recognized as a xenobiotic and "mistakenly" metabolized as an alternative substrate in tobacco cells. The function of the bio-transformed product was not characterized but was shown to possess antifungal activity (results not shown) which warrant further investigation.

The results of the current study as a collective provide an undisputed contribution to the metabolism of oximes in plants and also in the field of metabolomics. The use of UHPLC-MS in combination with chemometrics statistical-based models was shown to comprehensively contribute to understanding the biology significance behind the metabolic response induced by chemical inducer (INAP) and its metabolism in plants. To summarize the overall findings of the study, the general concluding remarks are presented (**Chapter 9**).

Aims/objectives of the study

- The main aim of the current study was (i) the initial experimental design and method development of UHPLC-MS components for optimal metabolomic studies in plant systems for downstream evaluation using such an analytical platform, (ii) to evaluate the effect of isonitrosoacetophenone on the metabolome of different plant systems and (iii) to establish the metabolic fate thereof with the aid of UHPLC-MS-based metabolomic investigations in conjunction with multivariate data models and other chemical intelligent software.

Chapter 2: Literature Review

2.1. Plant Defense Mechanisms

2.1.1. Summary

As sessile organisms, with no adaptive immune system, plants are always faced with a challenge of defending themselves against a wide spectrum of biotic and abiotic stressors. To defend themselves against these devastating factors, plants have developed a very complex defense mechanism, which is well researched in the context of plant-pathogen interactions. It is known that plants defend themselves against pathogen attack by activating a multi-component defense response, whereby a pathogen invasion is recognized by proteins encoded by receptor-like kinase (RLK) - and plant disease resistance (R) genes. The former bind/interact with microbe-associated molecular pattern molecules (MAMPs) while the latter bind/interact with specific pathogen-derived avirulence (Avr) effector proteins, resulting in the hypersensitive response (HR) that triggers host cell death in order to limit the spread of the pathogen (Chisholm *et al.*, 2006). The interaction between the microbe and plant defense proteins is well explained by the Zigzag model proposed by Jones and Dangl (2006) as illustrated by Figure 1.1.

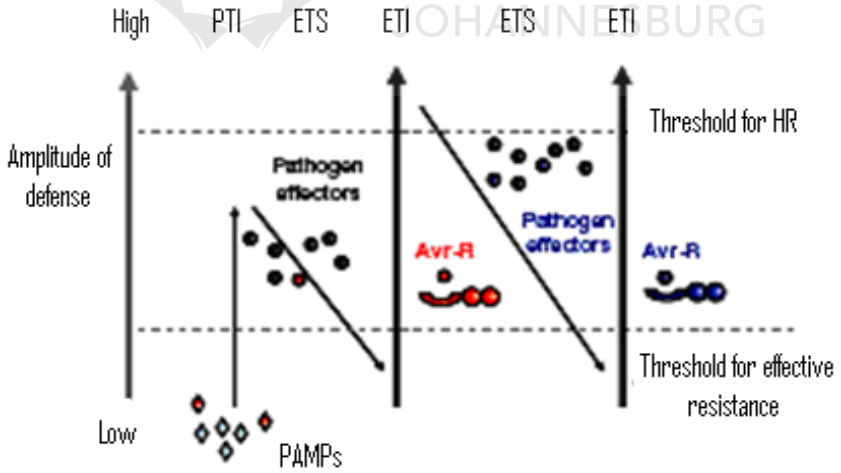


Figure 1.1: Schematic representation of the Zigzag model illustrating the underlying steps from gene-to-gene interaction which result in full implementation of plant defense response. PTI (PAMP-triggered immunity, ETS (effector-triggered susceptibility), and ETI (effector-triggered immunity) (Jones and Dangl, 2006).

From the above model (Figure 1.1), the ultimate amplitude of disease resistance or susceptibility is proportional to (PTI-ETS+ETI). The initial stage of this pathogen-plant interaction consist of the perception of microbial/pathogen-associated molecular patterns

(MAMPs / PAMPs, red diamonds) *via* pattern recognition receptors (PRRs) leading to a defense response known as PAMP-triggered immunity (PTI). In second and "intermediate" phase, successful pathogens, which have managed to overcome the PTI, release effectors proteins that interfere with PTI, and this subsequently leads to a state known as effector-triggered susceptibility (ETS). In the final phase of this model, one effector (indicated in red) is recognized by nucleotide binding leucine-rich repeats (NB-LRR) protein. This recognition initiates the activation of effector-triggered immunity (ETI), which is an amplified version of PTI that often passes a threshold for induction of hypersensitive cell death (Figure 1.1). It is also important to note that pathogen isolates that have lost the red effector are selected, and perhaps gained new effectors through horizontal gene flow (in blue), and this helps the pathogen to suppress ETI. Selection favors new plant NB-LRR alleles that can recognize one of the newly acquired effectors, resulting again in ETI (Tao *et al*., 2003; Jones and Dangl, 2006; Thilmony *et al*., 2006).

The HR response is then followed by a complex signaling network, involving cytosolic Ca^{2+} and H^{+} ions, reactive oxygen intermediates, as well as molecular signaling molecules (hormones) such as jasmonic acid (JA), salicylic acid (SA) and ethylene (ET) (Jones and Dangl, 2006), which together result in the induction of the defense mechanisms (McDowell and Dangl, 2000; Odjakova and Hadjiivanova, 2001). In another similar but independent case, plants do not produce the appropriate R proteins against Avr proteins of the pathogen. In such cases, the delayed cell response leads to the proliferation and spread of the bacteria to other parts of the plant and ultimately results in disease symptoms. This type of interaction is normally referred as a compatible interaction and commonly observed when host bacteria such as *Pseudomonas syringe* pv. *tabaci* infect a host plant (*Nicotiana tabacum*) (Huang *et al*., 1988; Jones and Dangl, 2006). Although the Zigzag model (Figure 1.1) is commonly used to show the fundamental basis of plant defense towards pathogens, it does not fully detail the underlying physiological changes which take place during plant defense responses. A more descriptive model (Figure 1.2) was also used in conjunction with the Zigzag model to fully gain the molecular detail of signal transduction processes which form the basis of plant defense.

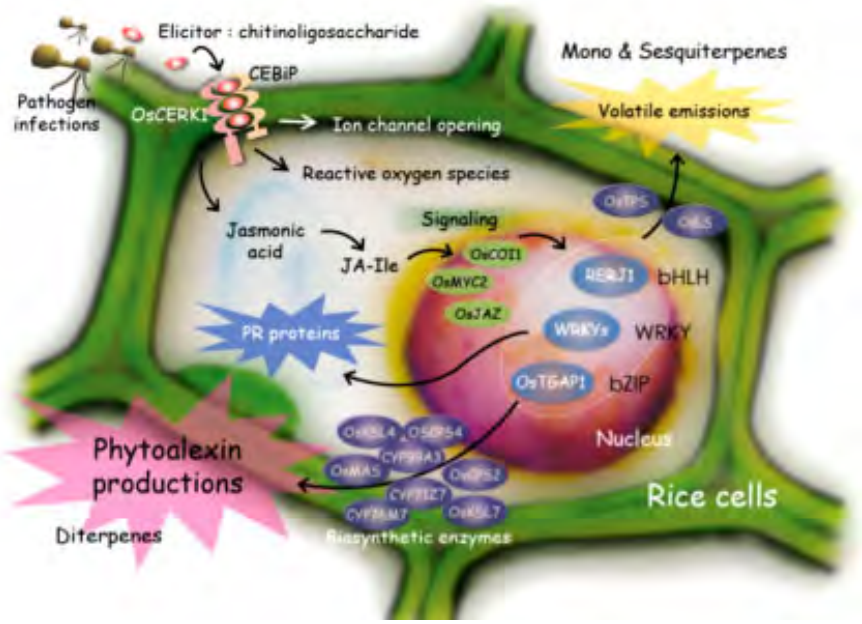


Figure 1.2: A schematic representation illustrating the main physiological pathways which takes place during plant-pathogen interactions using rice cells as a model (http://park.itc.u-tokyo.ac.jp)

2.1.2. Induced Resistance (IR): SAR and ISR

Plant innate immunity also includes an induced resistance (IR) aspect. By definition IR is a "physiological state of enhanced defensive capacity" elicited by specific stimuli, whereby the plant's innate immunity is potentiated against subsequent biotic challenges (Van Loon *et al*., 1998). IR is divided into different classes, depending on the mechanism in which they are implemented and exhibited (Pieterse *et al*., 2009). Using the model plant, *Arabidopsis thaliana*, the signaling pathways controlling these types of IR and defense responses in general are well characterized (Pieterse *et al*., 2009). The most common type IR is called the systemic acquired resistance (SAR); this phenomenon was previously known as physiological acquired immunity or just induced resistance, and is a physiological response in which plants exhibit a long-lasting response due to infection which subsequently results in a stronger response towards subsequent infection so that increased resistance or reduced disease symptoms are exhibited (Ryals *et al*., 1996).

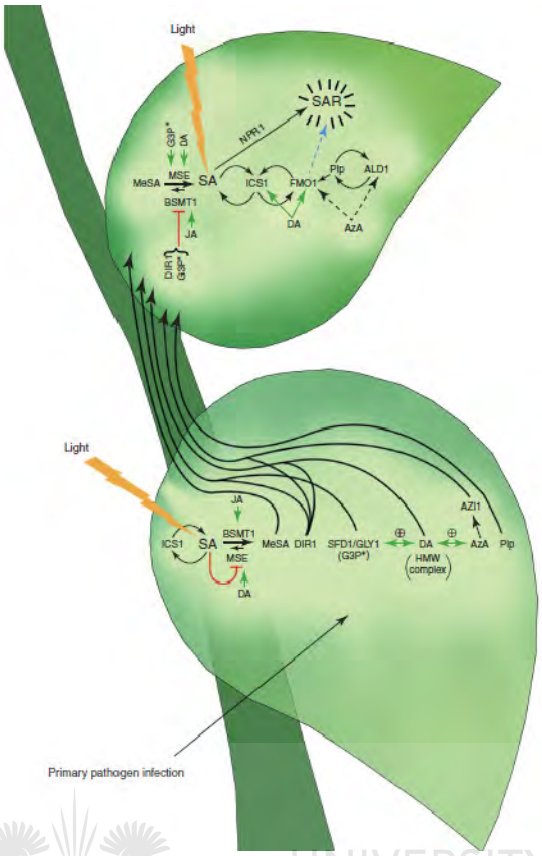


Figure 1.3: Schematic representation of out-lining different basic routes leading to the establishment of SAR.SAR signals are believed to be orchestrated by several components abbreviated on the figure such as ALD1 (AGD2-Like Defense response protein 1); AzA (Azelaic acid); BSMT1 (Benzoic acid/Salicylic acid Carboxyl Methyltransferase 1); DA (Dehydroabietinal); DIR1 (Defective in Induced Resistance 1); G3P (Glycerol-3-phosphate); FMO1 (Flavin-dependent monooxygenase 1); HMW (High molecular weight complex) ICS1 (Isochorismate Synthase); JA (Jasmonic acid); MeSA (Methyl salicylate); MSE (MeSA esterase); NPR1 (Non-Expressor of Pathogenesis-Related Genes 1); Pip (Pipecolic acid); SA (Salicylic acid); SFD1 (Suppressor of Fatty Acid Desaturase Deficiency 1). However, the mechanism through which some these signals interact is still unclear (Dempsey and Klessig, 2012).

The carrier of the message from the localized site of infection to the systemic sites is currently not fully understood, however, previous research has shown that hormones such as SA are believed to play significant roles in the host-pathogen interactions (Figure 1.3) (Zimmerli *et al*., 2000, 2001). Hormones such as SA, JA and ET play central roles during IR and depending on the type of hormones which are produced during IR, the outcome of resistance are different and as such different types of IR exist.

A different type of IR with similar characteristics to SAR has been well characterized in plants and in most cases it has been commonly mistaken for SAR itself, however, this type is known as induced systemic resistance (ISR). The difference between the two types of IR states is not physically obvious, but physiologically, these two forms of IR are completely different as they are activated differently and sometimes independently. In the context of plant defense responses, it is believed that SAR is triggered by biotrophic organisms whilst ISR is triggered by beneficial microorganisms known as plant growth promoting rhizobacteria (Ramamoorthy *et al*., 2001). In the latter case the phyto-hormones jasmonic acid/ethylene are believed to play a central role (Thomma *et al*., 1998; Glazebrook, 2005).

Although the two IR responses are physiologically different, it is important to note that there exists a cross-talk between respective phyto-hormone signalling pathways, which can be either antagonistic or synergistic and subsequently provide the plant with powerful regulatory potential (Spoel and Dong, 2008; Pieterse *et al*., 2009). This cross-talk is also believed to play roles in the enhancement of the ability of plants to use energy sparingly and create a generalized/flexible signaling network that allows a strong specific type of IR to be mounted against the invader (Van der Ent *et al*., 2008; Pieterse *et al*., 2009). The differences between the two types of IR responses are well outlined with the schematic representation below (Figure 1.4).

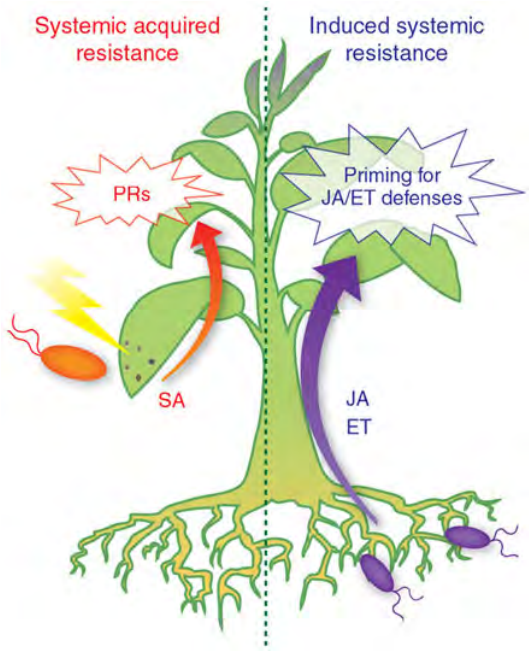


Figure 1.4: A pictorial comparison between SAR and ISR in plants. From the figure it can be seen that SAR is induced by abiotic or biotic elicitors, and it is dependent to SA and associated with the accumulation of pathogenesis-related (PR) proteins. On the other hand, ISR is induced by different strains of plant growth-promoting rhizobacteria (PGPR), and it is dependent on ethylene and jasmonic acid hormones. Though different in hormonal level, both responses are dependent on a functional NPR1 protein (non-expressor of pathogenesis-related protein) (Pieterse *et al*., 2009).

From the above information, it now clearer that these different forms of IR appear to be associated with direct activation / full implementation of plant defenses. However, another form of IR state which does not result in the full implementation of defenses has been shown, this IR state is known as priming. Plants that are under a primed state have been shown to have the ability to 'recall' a previous infection, however, this is only evident after secondary infection (Conrath *et al*., 2006; Goellner and Conrath, 2008). Full details regarding this complicated type of IR are discussed in subsequent sections.

2.1.3. Priming

As mentioned above, IR is a very complex phenomenon and currently not all information regarding the underlying physiological processes thereof is known. As seen from the previous section, IR is traditionally defined as either SAR or ISR. However, a new type of IR termed "priming" was described and is currently receiving significant research attention. To prime means generally to prepare or to make ready. On the other hand, in the context of plant

defense responses, priming is a physiological process by which a plant is predisposed to respond rapidly to future biotic or abiotic stress (Conrath *et al*., 2006; Goellner and Conrath, 2008). The condition of readiness achieved by priming has been termed the "primed state" (Conrath *et al*., 2006). The primed state may persist as a residual effect following an initial exposure to the stress. For example, the classical pathogen-induced HR is often induced with greater efficiency in plants that have previously experienced pathogen attack (Kuc, 1982). Priming therefore initiates a state of readiness that does not confer resistance *per se* but rather allows for accelerated induced resistance. One presumed benefit of priming is that it does not impose the costs associated with full implementation of a fully induced defense response (Frost *et al*., 2008).

In an attempt to conceptualize the primed state of plants, Frost and colleagues (2008) used a plant: herbivore interaction model, where there is an induced defense in plants in response to herbivore feeding. This defense response includes changes in a suite of chemicals that are toxic or unpalatable to herbivores. Furthermore, this induced resistance can be either indirect or direct. In the indirect case, the host produces volatile compounds, which are believed to function as attractants of natural enemies of the invading herbivore (Frost *et al*., 2008). Some of these released volatiles together with other signaling hormonal molecules such as JA are perceived by other undamaged parts of the host plant or even adjacent neighboring plants and induce a primed state in distant tissues or plants (Conrath *et al*., 2006). Although the full physiological mechanism of priming is still not fully known/ understood, priming is offering a great opportunity to researchers to gain knowledge about how to manipulate the plant defense responses, which will allow genetic manipulation resulting in plants that are resistant towards a wide spectrum of pathogens. There are no phenotypical or physical characteristics of the primed state, but recently some important factors which play central roles during priming have been shown. These include amongst many, the genes encoding proteins such as the non-expressor of PR genes 1 (NPR1) (Conrath *et al*., 2002), mitogen activated protein kinase (MAPK)-3 and -6 (Beckers *et al*., 2009), flavin dependant monooxygenase (FMO1) (Mishina and Zeir, 2006) and the lipid transfer protein (LTP) (Jung *et al*., 2009; Maldonado *et al*., 2002).

To date, several synthetic and naturally occurring molecules have also been used to induce the primed state against several bacterial and fungal pathogens. Such chemicals include riboflavin (Zhang *et al*., 2008), saccharin (Walters *et al*., 2009), sucrose (Gomez-Ariza *et al*.,

2007), hexanoic acid (Vicedo *et al*., 2009), beta-aminobutyric acid (BABA) (Ton *et al*., 2005), nicotinic acid (INA) (Gorlach *et al.,* 1996) and azealic acid (Jung *et al*., 2009). The biological mechanism of action of these chemical is not known, however, most studies are being carried out in attempt to fully characterize the mechanism of action of such chemical activators of defense. A good example is benzothiadiazole [benzo-(1,2,3)-thiadiazole-7-carbothioic acid S-methyl ester] (BTH), a functional analogue of the plant hormone-like molecule salicylic acid (SA) and the key player of SAR (Gorlach *et al*., 1996). It is one of the most researched molecules in the context of plant defense responses because of its ability to induce a more effective plant resistance response. This molecule has also been shown to prime the plant defense machinery against the pathogens (Friedrich *et al*., 1996; Gorlach *et al*., 1996). Similarly to SA, BTH efficiently inhibits hydrogen peroxide (H_2O_2) scavenger enzymes, ascorbate peroxidases (APXs) and catalases (CATs), but it is known to activate the NIM (non-induced immunity) / NPR1 (non-pathogenesis-related protein 1 inducer) genetic pathway (Wendehenne *et al*., 1998; Kohler *et al*., 2002). It is thus evident that BTH functions by modifying the redox homeostasis of the cells and results in full implementation of defense mechanisms which is orchestrated by H_2O_2, and activation of genes encoding pathogenesis related (PR) proteins and phytoalexins pathways (Gorlach *et al*., 1996; Faoro *et al*., 2008). BTH was also found to directly activate PR-1 and to prime *Arabidopsis thaliana* for potential phenylalanine ammonia lyase (PAL) expression in response to the infection by phytopathogenic *Pseudomonas syringae* pv. *tomato* (Pst) (Lawton *et al*., 1996). BTH's ability to induce PAL was also shown elsewhere (Maffi *et al*., 2011). In other separate but similar examples, BTH was also shown to induce SAR in tobacco (Friedrich *et al*., 1996) and in wheat (Gorlach *et al*., 1996). In an example which is more related to the current study, Dao and colleagues (2009), using NMR based metabolomics and gene expression studies, have also shown that BTH induce metabolic changes in *Arabidopsis* plants and also induced expression of some common PR proteins (Dao *et al*., 2009).

It is quite interesting to note that several promoters for genes encoding proteins functioning towards plant defense responses, such as that of PR-1, are known to respond to several chemical inducers (especially BTH) of defense (Gatz and Lenk, 1998; Padidam, 2003). It is thus convincible that the use of chemicals in the induction of IR is a promising area of research in the attempt to fully understand biological mechanisms involved in plant defense responses. This, in turn, will subsequently allow comprehensive genetic manipulation of plants to improve them to respond more efficiently against different pathogens.

Previous work on priming was mostly concerned on the same generation but only recently priming has been reported to be transgenerational (Pastor *et al*., 2012; Slaughter *et al*., 2012). Using *Arabidopsis thaliana* plants which were primed by BABA and two pathogens (bacteria and fungi), it was shown that the descendant of previously challenged plants were able to withstand subsequent treatment better than those originating from unchallenged/non-primed parental plants (Slaugheter *et al*., 2012). Together with other reports, the work of Slaughter and colleagues has shown that information underlying the molecular basis of priming is stored in the genetic make-up of primed plants and as such can be carried along to the next generation. A few examples on transgenerational priming were previously reported and these include wild radish that has been fed on by a pest (*Pieris rapae*) or treated with jasmonic acid mimicking herbivore feeding (Boyko *et al*., 2007). Progenies of virus-treated tobacco plants were also shown to exhibit stronger resistance (Boyko *et al*., 2010; Kathiria *et al*., 2010). The molecular basis of transgenerational priming is still unclear, however, for the offspring to remember a past experience from the parents, the latter need to perceive and store the information and subsequently transmit it to the descendants (Slaughter *et al*., 2012). For such to be possible, epigenetic mechanisms are believed to play central roles (Chinnusamy and Zhu, 2009; Alvarez *et al*., 2010; Sano, 2010), as well as histone modification (Boyko 2010; Luna *et al*., 2011) and small RNA molecules (Vaucheret, 2006).

Pastor *et al*. (2012) stated that the main mechanisms of priming are temporally dissected, however, the accumulation of ROS, callose, hormonal responses and other components of defense happen relatively early after challenge. However, the accumulation of unphosphorylated MAPKs, the modification of histones and DNA methylation are long lasting processes and the latter can be transferred to the offspring. The involvement of epigenetic modification during plant defense is not a new phenomenon (Alvarez-Venegas *et al*., 2007; Bezhani *et al*., 2007; March-Diaz *et al*., 2008; Berr *et al*., 2010).

2.1.4. Passive *versus* active defense

Apart from the above-presented inducible defenses (active), plants have an adjacent form of defense known as constitutive defense, which is constantly present. This constitutive defense (passive) is made up of different components which plants normally possess naturally for protection against different environmental threats.

Both active and passive defense mechanisms in plants are further divided into subclasses: physical and chemical barriers, which are pre-formed (for passive defense) or induced (for active defense). In the constitutive form of defense, plants rely on pre-existing anatomical barriers (such as trichomes, cuticles and cell walls) and pre-formed antimicrobial compounds to fight or stop the invasion of a pathogen or an abiotic stressor (Garcia-brugger *et al*., 2006). On the other hand, in the inducible defense, certain cellular mechanisms are actively unfolded to combat the stress. For instance, cell wall strengthening upon pathogen infection was shown to take place in the form of callose, lignin and suberin polymers appositions. Oxidative cross-linking of extensin proteins in the cell wall was also shown to take place (Garcia-brugger *et al*., 2006). Cell wall strengthening is a way of preventing further spreading to other unaffected parts of the plant. The expression of genes that encode proteins, believed to be involved in the cell wall strengthening phenomenon, has been previously shown to be induced in plants challenged with pathogen/pathogen-derived elicitors (Hammond-Kosack and Jones, 1996; De Ascensao and Dubery, 2000; Zwiegelaar and Dubery, 2006). Interestingly, studies have demonstrated that some abiotic stressors also induce the expression of such genes. For instance, heat shock/stress was shown to trigger the expression of genes encoding proteins that function towards cell wall formation (Yang *et al*., 2006).

2.1.5. Chemical defenses of plants

Looking at the chemical facet of plant defense, the same phenomenon of constitutive and inducible defense also exists. For an example, preformed chemicals exist and are known as phytoanticipins. These are antimicrobial compounds found in healthy/unchallenged plant tissues of which the function is to protect plants against natural pests (Van Etten *et al*., 1994). Plant metabolites in general provide protection against biotic or abiotic stresses (Dixon, 2001; Jahangir *et al*., 2009). Chemicals such as reactive oxygen species, phytoalexins, and other secondary metabolites are known to be part of inducible defense responses (Van Loon, 2000).

Research has shown that the accumulation of chemicals/metabolites (such as phytoalexins) due to stress, is an important component of the plant defense response. Phytoalexins are plant antimicrobial secondary metabolites, and they are synthesized from precursors such as phenylalanine, malonyl-CoA, acetyl-CoA, mevalonic acid and other amino acids (Hammerschmidt, 1999; Iriti and Faoro, 2009). During plant defense responses, phytoalexins are synthesized locally around infection sites, and the systemic accumulation thereof has never been shown (Van Loon, 2000). Apart from phytoalexins, other classes of compounds that have been shown to participate in plant defense responses include, amongst many, fatty acids, phenylpropanoids and flavanoids (Bollina *et al*., 2011). Some of these plant compounds (physiologically known as secondary metabolites) accumulate as stable molecules and some exist as precursors. For instance, activation of the phenylpropanoid pathway compounds is believed to result in the synthesis of a remarkably vast array of low-molecular-mass natural chemicals, which further act/participate as substrates in many cellular regulatory processes which includes plant defense responses (Jahangir *et al.,* 2009). In the context of pathogen- inducible secondary metabolites, camalexin represent a well studied phytoalexin (Beets *et al*., 2012). This molecule is produced by the model plant, *A. thaliana*, as well as other closely related species, and its involvement in plant defense responses has been shown and characterized (Thomma *et al*., 1999). The fact that this molecule exists in a specific group of related plants is believed to suggest that camalexin represents a relatively novel evolutionary invention in the competition between plants and pathogens (Bednarek *et al.,* 2011; Bednarek, 2012). The study of such and other related compounds in plants has recently received much attention. Metabolite accumulation in plants during pathogen attack is also finding increasing attention as a way of evaluating the effect of plant: pathogen interactions, for instance in plant: bacteria interaction (Hagemeier *et al.,* 2001; Tan *et al.,* 2004), plant: virus (Choi *et al.,* 2006) or plant: fungal (Wolski *et al.,* 2010). It is thus convincible that studies involving dynamic changes of metabolites in plant tissue undergoing stress responses represent a promising field of study in plant biotechnology. The involvement of plant metabolites in defense responses will also be discussed comprehensively throughout different sections of this manuscript.

2.2. Plant Secondary Metabolites

2.2.1. Overview

Plants produce a wide spectrum of compounds or metabolites, many of them genus– or species-specific. The plant metabolome is thus defined as the total collection of metabolites (low molecular weight compounds) that are present in cells or organisms and that participate in metabolic reactions required for growth, maintenance and normal function (Oliver *et al.,* 1998; Beecher, 2004). Based on the function, metabolites are normally referred to as either primary or secondary metabolites. Primary metabolites (nucleotides, amino acids, lipids and sugars) are known to function in simple maintenance of the cell such as growth and development (Verpoorte, 2000). On the other hand, secondary metabolites are compounds of which biosynthesis is restricted in a certain taxonomical group (Pichersky and Gang, 2000) and the functions are mainly interaction between the organism and its environment, ensuring the organism's survival in an ecosystem (Verpoorte, 2000). Plant secondary metabolites (also known as natural products) have been studied intensively. The importance of these compounds varies according to the respective activities thereof within the organism. The ability of an organism to produce secondary metabolites is associated with different reasons, for example, floral scent volatiles and pigments serves to attract insect pollinators for reproduction purposes (Dudareva and Pichersky, 2000).

As previously pointed out, plants synthesize secondary metabolite chemicals as a way of defending themselves against pathogens and herbivores but also to out-compete the growth of neighboring plants (Bennet and Wallsgrove, 1994; Dixon *et al*., 1996; Harborne, 1999). The involvement of plant metabolites in different plant developmental and physiological stages is well summarized by Figure1.5 which shows different domains in which these metabolites are deployed by the plants as a means to adapt well in given surrounding environmental conditions. By virtue of being the end product of cellular processes, metabolite constituents of the cell at any given time determine the phenotypic trait of an organism (Fernie *et al*., 2004). In recent years the studies of plant metabolites have received overwhelming attention. For example, in the biotechnology field, metabolic pathways are being targeted for genetic engineering (Dawson *et al.,* 1989; Bak *et al*., 2000). Manipulation of such pathways and knowledge gained thereof is of major importance as it opens up ways for genetic engineering,

aiming to increase resistance against microbial infections or insects attack (Dawson *et al* 1989; Dixon and Paiva, 1995).

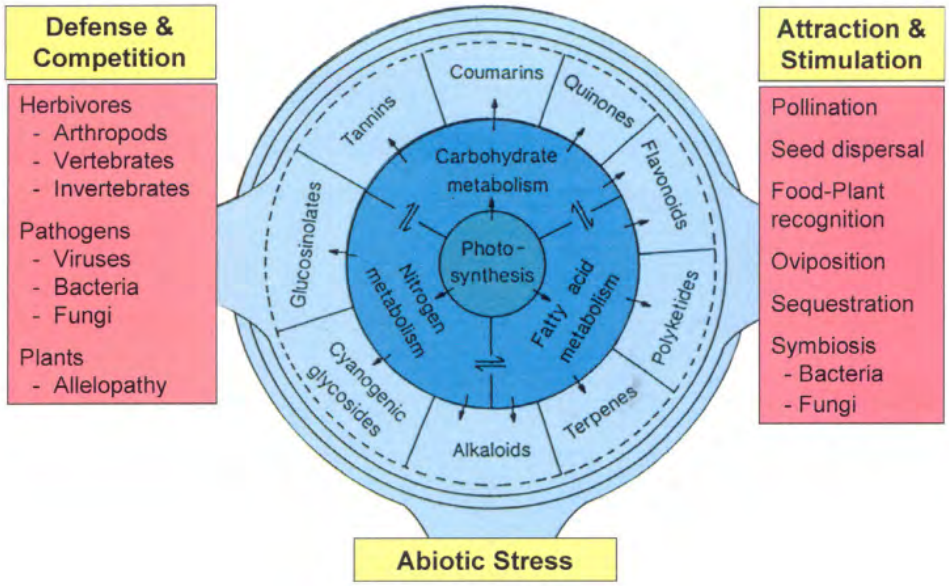


Figure 1.5: A schematic representation of some known ecological functions of plant-derived secondary metabolites (Hartmann, 2007).

According to Verpoorte (2000), plant secondary metabolites can be classified based on determined chemical features, biosynthetic origin or plant origin. There are three major pathways which most secondary metabolites originate from. These pathways are namely the shikimate, isoprenoid and polyketide pathways. In most cases secondary metabolites are mainly classified based on structural skeletal properties which is unique for certain biosynthetic pathways. Thus metabolites from similar pathways are more likely to have similar structural features, and a very good example will be terpenoids which are produced from isoprenoid pathway using C5 building block to build up C10 (monoterpenes), C15 (sesquiterpenes), C20 (diterpenes), C30 (steroids and triterpenes) and C40 (carotenoids) compounds (Verpoorte, 2000). The building blocks for phenylpropanoids are either from phenylalanine or tyrosine (C9) and on the other hand, polyketides are produced from acetate (C2) (Verpoorte, 2000). Overall, plant biosynthetic pathways are very complex systems that comprise different sub-pathways with different functions. For the scope of the current study, only two pathways will be considered, namely the shikimate pathway (Figure.1.6) and the phenylpropanoid pathway (Figure.1.7) as they relate to topic of the thesis.

2.2.2. The Shikimate Pathway

This pathway is exclusively found in microorganisms and plants, and it is also known to be the major source of most aromatic compounds (Bentley, 1990, Haslam, 1993; Hermann, 1995; Schmidt and Amrhein, 1995). The major steps which are involved in this pathway are summarized in Figure. 1.6. Other than shikimate itself, chorismate is also formed which is further converted to phenylalanine, tyrosine and tryptophan. Other metabolites such isochorismate, 4-hydroxybenzoic acid and 4-aminobenzoic acid also exist as precursors from which a series of different secondary metabolites are formed (Verpoorte, 2000). It is believed that two versions of the shikimate pathway exist in plants, one which yields aromatic amino acids (Jensen, 1986; Herrmann and Weaver, 1995) and the other secondary metabolites. However, there is no sufficient evidence so far to support this hypothetical theory. It is also believed that genes encoding enzymes which are parts of the shikimate pathway originate from at least three ancestors (cyanobacteria) (Vogt, 2010).

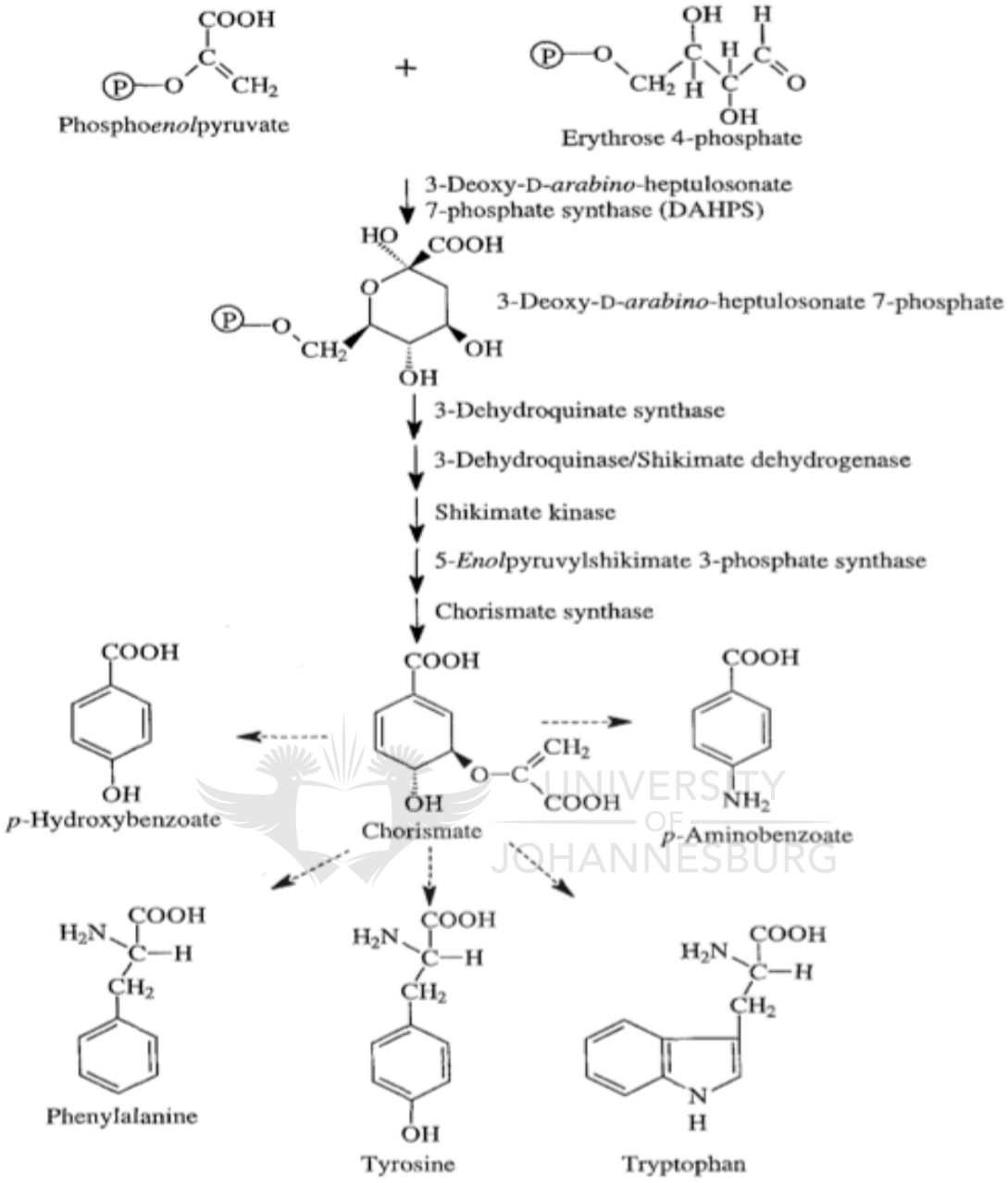


Figure 1.6: The shikimate pathway of higher plants (adapted from Herrmann and Weaver, 1999).

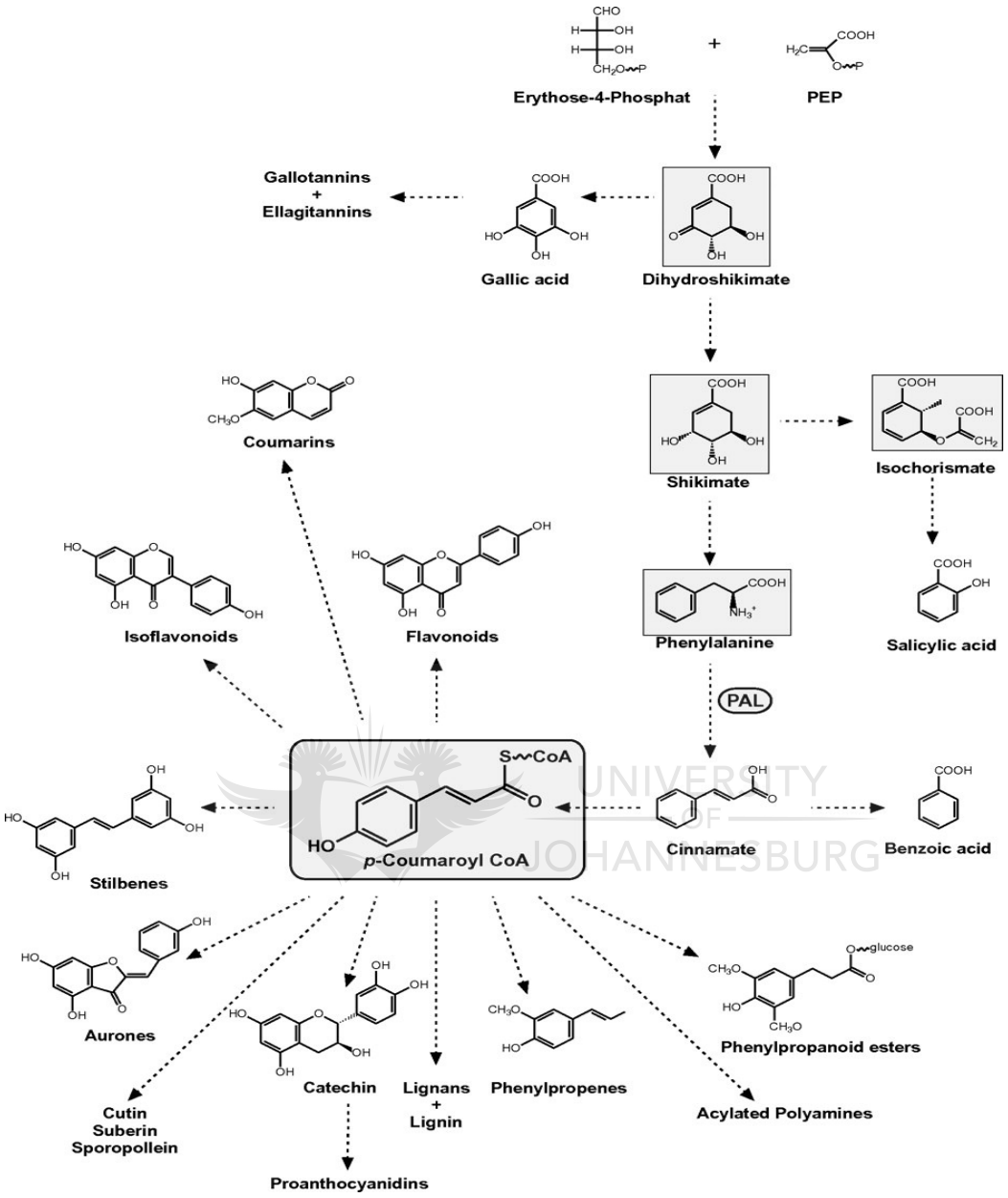


Figure 1.7: Schematic representation of important branching points of the phenylpropanoid pathway which results in the production of a variety of different secondary metabolite classes. Metabolites which are derived from the shikimate pathway are shaded in gray (Vogt, 2010).

2.2.3. The Phenylpropanoid Pathway

The phenylpropanoid pathway should be thought of as the continuation of the shikimate pathway, as most of the shikimate pathway products are channeled into this pathway *via* phenylalanine. Just as in the shikimate pathway, the phenylpropanoid pathway yields a wide array of secondary metabolites (Vogt, 2010). The main metabolites from this pathway are lignin and flavonoids. However, the phenylpropanoid pathway also results in other secondary metabolites such as coumarins, phenolic volatiles and hydrolyzable tannins (Figure.1.7) (Vogt, 2010).

In the context of plant defense, this pathway is acted upon by different stimuli ranging from biotic to abiotic (La Camera *et al*., 2004). The resulting polymers which are synthesized from this pathway such as lignin are known to play a vital role during cell wall strengthening (see section 1.1.4). There are several genes and enzymes which participate in this pathway and function collectively even though they catalyze different steps. The phenylpropanoid pathway contains three mandatory steps catalysed by Phenyl Ammonium Lyase (PAL), cinnamate 4-hydroxylase and 4-coumaroyl CoA ligase. All these enzymes yield "intermediate metabolites" which provide the basis for the synthesis of some important subsequent metabolites. The link between the shikimate- and phenylpropanoid pathway is a very crucial one and it is only possible under strict/controlled physiological conditions. It is important to note that under sub-optimal conditions such as low photosynthetic conditions (i.e. when plants are under stress), low concentrations of shikimate intermediates are channeled into the phenylpropanoid pathway, redirecting the latter towards the production of phytoalexins, volatiles, flavanoids, anthocyanins and production of defense related proteins (Abdulrazzak *et al*., 2006; Schoch *et al*., 2006). Soluble phenyls from this pathway, with defense-related and antioxidant activities such as caffeoyl quinate (chlorogenic acid) and caffeoyl phenyllactic acid (rosmarinic acid) are predominately present in the Solanaceae and Lamiales genera respectively (Petersen *et al*., 2009) The diversity of enzymes comprising this pathway makes it very interesting, because this is the reason why this pathway yields such a wide variety of secondary metabolites. The diversity of metabolites resulting from this pathway is due to wide variety of enzymes involved. Some of these enzymes exhibit limited specificity; whilst others are very specific to certain reactions. It is also remarkably to note that there are more than 300 functional genes encoding cytochrome

P450s (CYP450) in *Arabidopsis* of which only one of them encode the hydroxylation of trans-cinnamate to 4-coumarate. The mutation on these CYP450 genes has been shown to have an effect on the growth and development of *Arabidopsis* plants, and other than growth and development impediments, accumulation of unusual cinnamoyl malate products were also observed in transgenic plants plants (Schilmiller *et al*., 2009) highlighting the importance of each and every enzyme participating in this pathway. In lignin synthesis which is the main final step of this pathway, hydroxylation and methylation steps (carried out by CYP450 and O-methyltransferase, respectively) are important during stress-related responses (Dohleman and Long, 2009). The phenylpropanoid pathway is thus one of the most important pathways that yield essential secondary metabolites involved in stress tolerance, growth and developments of plants. The various structural decorations on substrate metabolites result in a wide spectrum of metabolites and subsequently in a large and complex metabolome (Vogt, 2010) (Figure. 1.8).

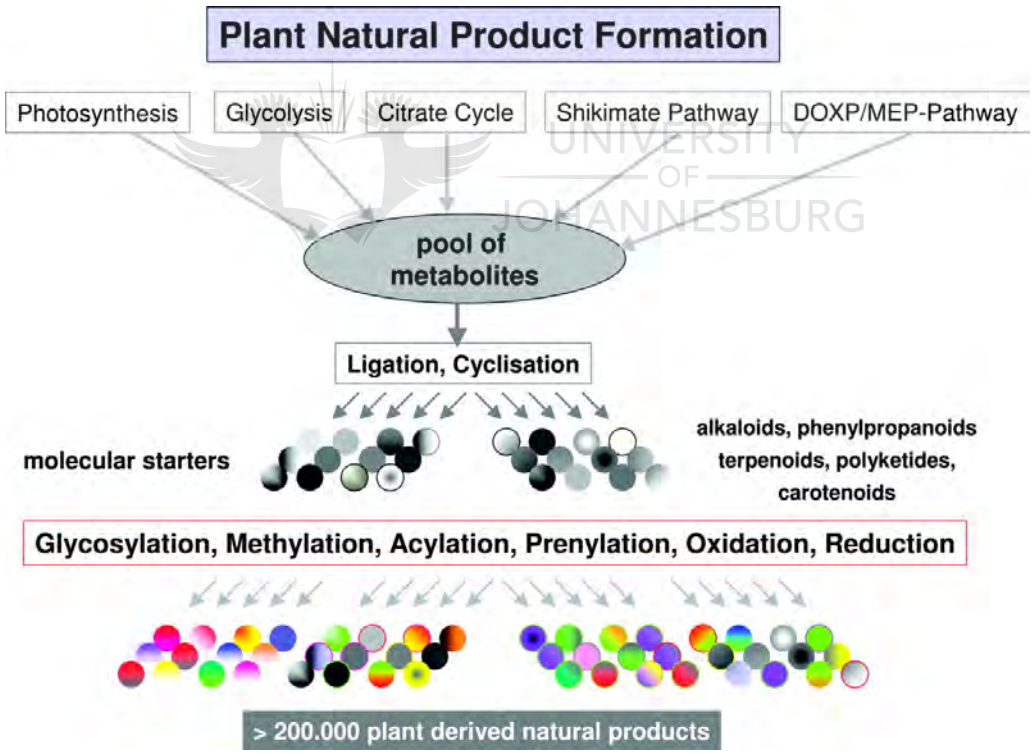


Figure 1.8: Schematic representation showing different convergence and divergence points during the biosynthetic pathways of secondary metabolites. The diversity of secondary metabolites is due to different chemical "decorations" as indicated (Vogt, 2010).

2.2.4. Cyanogenic Glycosides and Glucosinolates

Cyanogenic glycosides and glucosinolates are very specialized classes of molecules which form part of plant defense responses, specifically against insects/herbivore attack (Bak *et al*; 2000; Moller, 2010). The presence of these compounds is limited to certain plant species such as *Sorghum bicolor* and *Arabidopsis thaliana* (Bak *et al*., 2000), and the biosynthetic pathways and respective involvement in plant defense responses are well characterized (Bak *et al*; 2000; Jørgensen *et al*; 2005). It was also further shown that biosynthesis thereof is a well controlled system and enzymes which are involved during these biosynthetic pathways are arranged in a specialized unit called a metabolon (Jørgensen *et al*; 2005; Nielsen *et al*; 2008**;** Moller, 2010). This arrangement of enzymes is thought to be a strategic plan aimed at efficient production of these labile/unstable molecules. Looking at the specific biosynthetic pathways (Figure 1.9), it can be noted that there exists multiple precursors which, if not rapidly converted would be broken down, (e.g. *p*-hydroxymandelonitrile, an intermediate during cyanogenic glycoside (dhurrin) biosynthesis). If not rapidly glycosylated by the UGT85B1 enzyme it can easily dissociate into hydrogen cyanide and *p*-hydroxybenzaldehyde (Jørgensen *et al*; 2005) and as such the main product, dhurrin will not form. With the aid of gene fusion technology, the location of these biosynthetic enzymes was monitored using spectral variants of green fluorescent protein by means of confocal laser scanning microscopy and it was further demonstrated that these enzymes formed distinct domains (metabolons) in the endoplasmic reticulum (ER) (Koch *et al*., 1995; Bak *et al*., 1998; Winkel, 2004). Once cyanogenic glycosides and glucosinolates molecules are formed, they are stored and only become active when a host plant is under attack. During this stage, these molecules are converted into toxic products through a process called the 'cyanide -' or 'mustard bomb', respectively (Osbourn, 1996). For instance, glucosinolates are activated by the plant enzyme myrosinase (a thioglucosidase) and as a consequences they break down into variety of products, including isothiocyanates, nitriles, and thiocyanates, all of which are highly reactive compounds (Osbourn, 1996). The involvement of cyanogenic glycosides and glucosinolates are well described and discussed throughout several sections of this thesis.

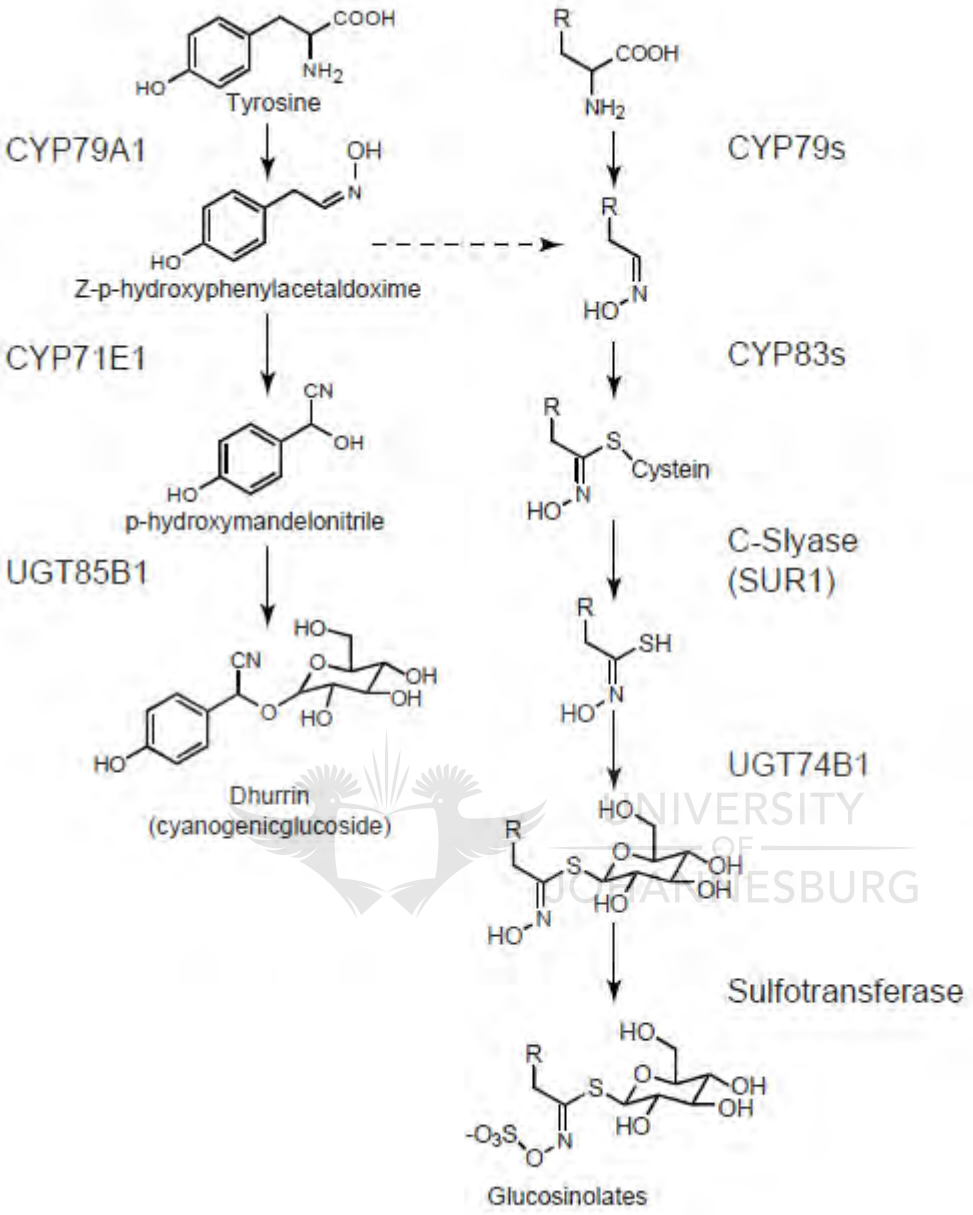


Figure 1.9: Schematic representation showing the biosynthesis pathways of cyanogenic glycoside and glucosinolates in plants. The metabolic crosstalk is shown in dashed arrows (Jørgensen *et al*., 2005).

2.3. Metabolomics

2.3.1. Overview

From the central dogma of molecular biology, cellular information flows from DNA which is transcribed into messenger RNA (mRNA). The latter is also subsequently translated into proteins, and some of these proteins participate in biosynthetic pathways that lead to the production of metabolites (Figure 1.10). Metabolites, as small molecules that participate in general metabolic reactions are required for the maintenance, growth and normal function of a cell (Hanhineva, 2008). The flow of the information, as stipulated by the central dogma, is well coordinated, and the components of this dogma are interdependent. It is therefore important to note that the flow of information is not unidirectional; there are several back-loops relaying information in multidirectional manner. For example, some proteins are known to function as transcription factors and hence activate DNA transcription, and in another example some RNA molecules are also known to interact with DNA as either regulators or suppressors of DNA-transcription, a classical example of such phenomenon could be that of small interfering RNA (siRNA). It is through these interactions amongst the members of the central dogma components, that a cell acquires its full functionality of its cellular metabolism. By virtue of being the end product of most cellular metabolism, metabolites act as regulatory components of metabolism and hence accumulation thereof allow a complete 'snapshot' of the physiological status of the cell (Fernie *et al*., 2004; Ryan and Robards, 2006; Seger and Sturm, 2007).

When integrated with other "-omic" techniques, measurement of metabolic changes has lead to the identification of so called "silent phenotypes", genes whose expression does not affect the physical characteristics or behavior of an organism (Griffin *et al*., 2002). The metabolite content of the cell can be used to evaluate the physiological/phenotypic characteristics of the cell at given time (Lindon *et al*., 2003). Thus, a new field of study, termed metabolomics, was recently developed. By definition, metabolomics is an unbiased approach aimed at measuring the metabolite content of the cell, tissue or organism under given physiological status (Nicholson *et al*., 1998; Oliver *et al*., 1998). Table 1 shows metabolomic strategies that have been developed

for the study of metabolome. Most of these terminologies (in Table 1) are sometimes used interchangeably.

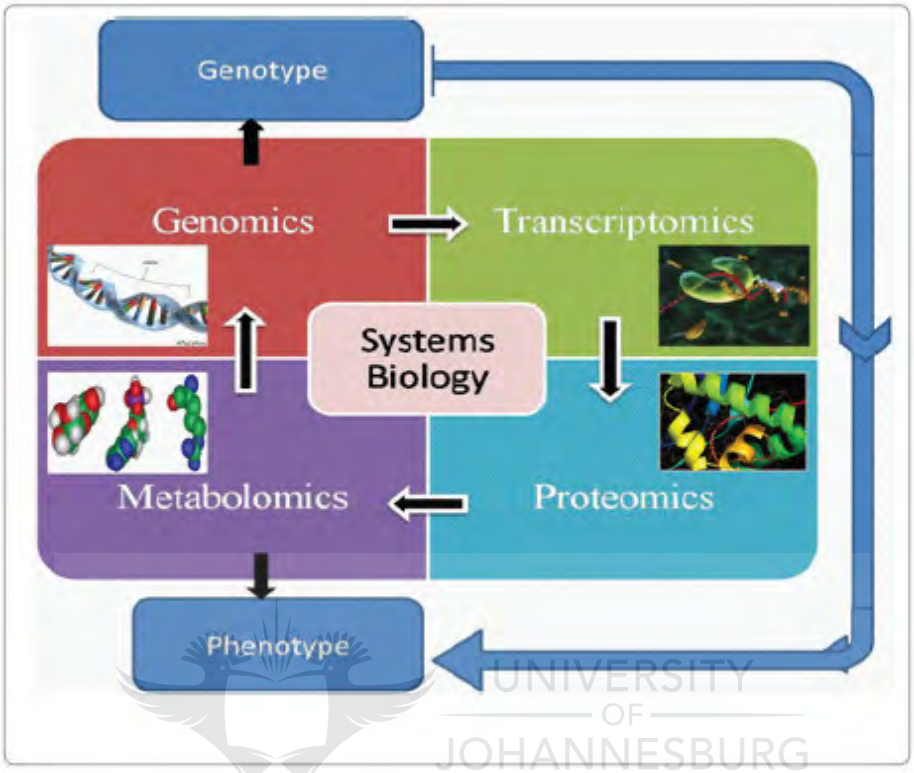


Figure 1.10: The schematic representation of the flow of cellular information which comprises the systems biology of the cell. The arrows indicate the direction in which information flows from/to one component of the dogma to another (Adapted from Nanda *et al*., 2011).

From Table 1 it can be seen that the term metabolomics can also be used interchangeably with metabonomics and even though they have different definitions, the meaning of the definitions are quiet similar and they basically differ based on sample nature (Lindon and Nicholson, 2008). Depending on the merit of the application, other terms can also be used and, as such, the use has also become individual preference. Irrespective of the term used, all these metabolomic approaches apply holistic analytical strategies to collect either targeted or untargeted analytical data to investigate the metabolites' distribution patterns and concentrations in bio-fluids, cells or organs of an organism under study (Theodoridis *et al*., 2011).

Table 1: Description of terminologies which are commonly used in conjunction or interchangeably with the term metabolomics. The compressive explanation of metabolomics as given in text is highlighted in red (Adapted from Hanhineva, 2008)

Terminology	Definition
Metabolome	The complete set of metabolites in a cell, tissue or organism
Metabolomics	Unbiased identification and quantification of all metabolites in a biological system, or a complete set of metabolites in a cell or tissue type
Metabonomics	A term not used in plant science, although sometimes used interchangeably with metabolomics. Mostly used in disease diagnostics and toxicology as the quantitative measurement of the dynamic response of living systems to pathophysiological stimuli or genetic modification
Metabolite (or metabolic) profiling	Quantitative analysis of a set of metabolites in a selected biochemical pathway or a specific class of compounds; biased/targeted approach which addresses only limited number of often structurally closely related analytes due to methodological limitations of the analytical platforms used
Metabolic fingerprinting	High-throughput, rapid global analysis; unbiased global screening approach to classify samples based on metabolite patterns or "fingerprints" that change in response to disease, environmental, or genetic perturbations, not necessarily involving analyte identification and quantification
Metabolic footprinting	Fingerprinting analysis of extracellular metabolites in cell culture medium as a reflection of metabolite excretion or uptake by cells

Furthermore, new fields of study have also recently been developed, of which the origin can easily be traced from metabolomics and can be thought of as the expansion of applications of metabolomics. One of these new fields of study is known as pharmacometabolomics. The latter is an approach aimed at investigating the effect of pharmaceutical and naturally occurring drugs on certain diseases/physiological disorders (Wang *et al.,* 2012). This approach represents a very promising prospect of metabolomics, because one area of considerable interest in the field of metabolomics is its application in the field of drug discovery and development, where the analysis of biofluids and tissues can provide a global view on the changes and distribution patterns of endogenous metabolites due to cellular responses to perturbations of drug treatment (Chan *et al.,* 2011; Kinross *et al.,* 2011). Furthermore, metabolomics can be useful and impact in several points in the drug development chain. This can be in target identification, lead discovery and optimization, preclinical efficacy and safety assessment, mode-of-action and mechanistic toxicology and pharmacological monitoring (Wang *et al*., 2012).

Another term which is partially stemmed from the field of metabolomics is that which deals with drug metabolism by living organisms. In their study, Holmes *et al*. (2007) used the term "xenometabolome", and they described it as the multivariate signature/description of the xenobiotic (foreign compound) metabolite profile of an individual or sample from an individual that has been exposed through any route (either deliberately or accidentally) to drugs, environmental pollutants, or dietary components that cannot be completely catabolized by endogenous metabolic enzyme systems (Holmes *et al*., 2007). In their approach, Holmes and colleagues used the combination of statistical methods and proton NMR-based metabolomics to characterize structural pathway connectivities of metabolites of commonly used drugs in urine of human beings. In this study, it was shown that statistical connectivities between drug metabolites can be established in routine "high-throughput" NMR screening of participants' samples that have randomly self administered drugs prior to the analyses. It was also shown that this approach should be of value in considering inter-population patterns of drug metabolism in epidemiological and pharmacogenetic studies (Holmes *et al*., 2007).

Though the above two examples serve as good platforms in which metabolomics was applied, there are also other fields in which metabolomics can be employed. For instance, metabolomics

has played significant roles in many fields such as responses to environmental stress (Lin *et al*., 2006; Viant, 2007), cancer research (Pyo *et al*., 2008), comparing different growth stages, physiological disorders (diabetes for an example), microbial studies (Koek *et al*., 2006; Khoo and Al-Rubeai, 2007; Mashego *et al*., 2007), studying global effects of genetic manipulation, nutrition and health (Van der Greef *et al*., 2004; German *et al*.,2005; Fava *et al*., 2006; Goodacre, 2007), natural product discovery (Wang *et al*., 2011a), in drug efficacy/toxicology and disease prognosis (Ackerman *et al*., 2006; Dieterle *et al*., 2006; Kell, 2006; Lindon *et al*., 2006; Chen *et al*., 2007; Homes *et al*., 2007; Van der Greef *et al*., 2007) and plant studies (Kopka *et al*., 2004; Weckwerth and Morgenthal, 2005; Hall, 2006; Kim *et al*., 2010).

To a certain extent, it is only lately that the comprehensive understanding of metabolomics has come to the fore. However, the origin of this field is well documented by Gates *et al*., (1978):

- In the late 1940s some scientists started to investigate the metabolite distribution patterns to make sense of some physiological disorders. Example includes studies which the normality issue, profiling body fluids and examining individual changes, inter-individual changes and later pathological conditions such as those found with alcoholics and schizophrenic patients were investigated (Van der Greef and Smilde, 2005).
- The growth of this type of study was quite noticeable in the 1960s when the technology was rapidly increasing and several groups in clinical chemistry started to take note of the inborn errors of metabolism using GC and GC-MS and this is well documented by (Politzer *et al*., 1976; Jellum, 1977; Jellum, 2001).
- However, it is during the 1970s where the combination between pattern recognition methods and analytical platforms has been vastly utilized for metabolite analyses. The work performed at the Fundamental Research on Matter (FOM) Institute, where profiling by pyrolysis mass spectrometry of complex biological samples has been performed, serves as a good example (Meuzelaar *et al*., 1973). It is worth noting that a scientist by the name of Jan Van der Greef stands as one of the individuals who first introduced the notion/concept of combining the mass spectrometric - and pattern recognition techniques as we experience them today.
- In 1983, Van der Greef, presented his findings to the audience of the first international chemometric meeting in Petten in the Netherlands. In his work, differences between

females and males were shown by principal component analysis (PCA) of the data of urine metabolites which were prior analyzed by field desorption mass spectrometry and subsequent analyzed by PCA (Van der Greef *et al*., 1983).

- It is also during this time that the interface between liquid chromatography and mass spectrometry has been shown for the analysis of plant metabolites (Games *et al*., 1984).
- Tas and Van der Greef (1995) reviewed on the improvements of the technology, especially on direct chemical ionization in combination with pyrolysis based mass spectrometry, and combination of the latter with pattern recognition models became the norm during the 1980s where it was mainly used for disease diagnosis, etiology and reverse pharmacology.
- Other techniques such as NMR have also found application in the analysis of complex biological samples. The work of Nicholson and Wilson (1989) stands out amongst other as the pioneering beginning if not the first in which profiling of bio-fluids by NMR is utilized (Bell *et al*., 1989; Nicholson and Wilson, 1989). The earlier work in which the interface between NMR and pattern recognition became active is also reviewed by (Nicholson and Wilson, 1989).
- Today, the analysis of metabolites has become much involved due to advanced instrumentation and incorporation of pattern recognition models which ease the analyses of the data.

By virtue of being amongst the fastest growing fields of systems biology, metabolomics result in highly complex data. As such, scientists in the field of metabolomics spend a great deal of time and effort extracting meaningful information from such multi-complex datasets (Robertson, 2005). The design of a metabolomics experiment is very crucial so as to lead to valid and reproducible results with meaningful biological knowledge/insights. Unlike other –omics approaches, metabolomics is a data-driven, or inductive, scientific discipline compared to the hypothesis-driven strategies employed in traditional molecular biology (Kell and Oliver, 2004). In its simplest form, metabolomics data sets can contain measurements performed on subjects under different biological conditions. This then results in significant complex data matrices which are cumbersome to analyze manually or with traditional statistical tools. As mentioned before, exploratory data mining methods are then the preferred choice for such analyses. As seen

from most examples herein the manuscripts, techniques such as PCA partial least square discriminate analysis (PLS-DA) and orthogonal projections to latent structures discriminate analysis (OPLS-DA) are commonly preferred for such involved analyses.

Like all other scientific experiments, there are different steps which are mandatory to follow when metabolomics experiments are conducted. All steps of the experiment should be appropriately validated and experimental aspects such as sample numbers, sample preparation, choice of analytical technology and data processing strategies are of significant importance (Broadhurst and Kell, 2006). Similar to transcriptomics (Ball and Brazma, 2006) and proteomics (Taylor *et al*., 2006), there exist accurate reporting standards for metabolomics data and should also be followed and these are currently available (Plumb *et al*., 2006). Figure 1.11 highlights a general overview of steps which are mandatory for a successful metabolomics study. These standards have been developed by the metabolomics standards initiative (MSI) to enable other researchers to document research accurately (Fiehn *et al*., 2007). It is of major importance to note that these guidelines do not describe how to perform investigations but rather gives a description as to how the experimental work should be carried out and this will, in turn, allow others to evaluate or repeat the work. This will also ensure that all possibilities of bias reporting are eliminated (Ioannidis, 2005; Ransohoff, 2005). However, there exists certain factors which are not easily removable such as those due to the age differences between the samples (Yu *et al*., 2007) and some stringent factors such as those due to diurnal variations (Plumb *et al*., 2005). The standards reported by Fiehn *et al*. (2007) were also followed during the progress of the current study. As stipulated above, the subsequent sections (2.3.2- 2.3.4) of this chapter will in most cases outline the basic mandatory steps of the metabolomics study.

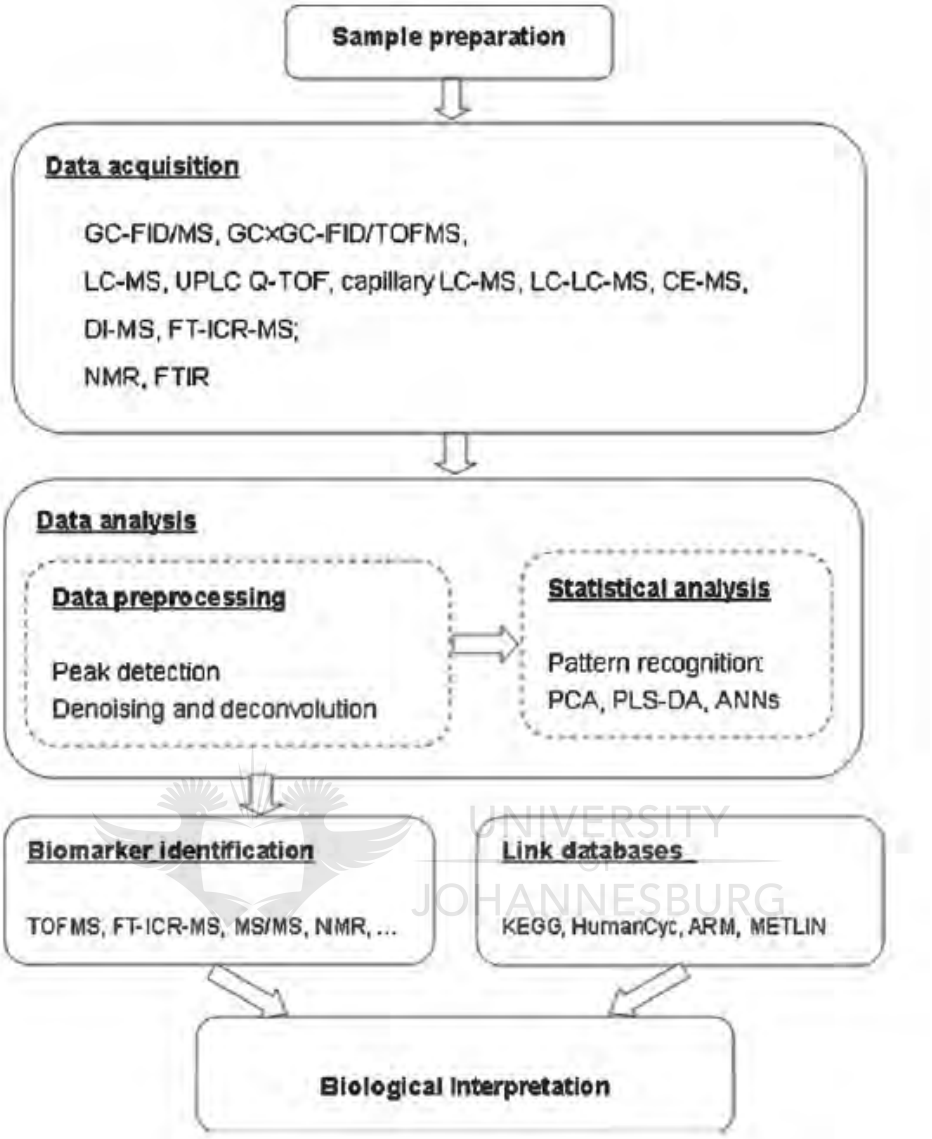


Figure 1.11: Schematic representation/overview of necessary steps which are commonly mandatory for most comprehensive metabolomics studies (adapted from Lu *et al*., 2008).

2.3.2. Sample Preparation

"The quality of samples and handling determines the quality of outcomes of the metabolites analyses" (t'Kindt *et al.,* 2009; Alvarez-Sanchez *et al*., 2010; Kim and Verpoorte, 2010). It is therefore important, to start with a good sample for better outcomes. Different specimens can be used from plants for metabolite analysis, and these can be in form of leaves, stems, roots and

fruits, etc. Different tissues will give different metabolic pictures, for example it was found that metabolite patterns of leaf veins are different from those of other parts of the leaf (Choi *et al*., 2004a&b). Another important aspect to take into account is the harvesting of samples, because it is in this step where undesirable chemical and enzymatic reactions of metabolites can occur and as a consequence, wrong conclusions can be made (Maltese *et al.,* 2009). It is therefore recommended that during harvesting, samples be kept in a cool environment in order to overcome the above-mentioned problem (Kim *et al*., 2010). The process of stopping the physiological processes (enzyme action / metabolism) is referred to as quenching. Different methods to achieve quenching have been proposed, drying of samples is one of them as excess water can also provide a suitable environment for enzymatic reactions (Kim *et al*., 2010). Freeze-drying the samples before analysis have been applied to remove excess water, however, this process also possesses some problems as it was found to lead to loss of metabolites, especially volatiles (Hanhineva, 2008).

Other than sample handling, there are other factors which can influence the quality of the analysis and these include the time of harvesting, age and gender (Plumb *et al*., 2005; Laiakis *et al*., 2010). With the aid of UHPLC-MS and multivariate data analysis (MVDA), Plumb *et al*. (2005) reported diurnal-dependant metabolite variation in mouse samples which was primarily caused by different sampling times. Samples collected at different times of the day (morning, afternoon or evening) were found to contain varying metabolites (or metabolic signatures/patterns). The levels of some primary metabolites such as sugars, malic acid and citric acid were also found to differ from one sample to another throughout the day cycle (Queiroz, 1974, Kim *et al*., 2006). This latter example illustrates that consistency in sample collection and preparation is mandatory for maximum data output in metabolomic studies.

One of the holy grails of metabolomics is to establish a “Universal” sample preparation technique, thus a single solvent metabolite extraction protocol which can enable isolation of all types of metabolites (Choi *et al.,* 2011). Due to the complexity of the physico-chemical properties of metabolites and respective matrix, it is very unlikely that there is a single solvent which can extract all metabolites from a cell or tissue under investigation (Kim *et al*., 2010). For NMR based metabolomics, perchloric acid was found to be a more suitable extraction protocol

for polar compounds such as sugars and amino acids (Kruger *et al*., 2007, 2008) but this method is also known to cause problems to some metabolites which are acid-intolerant. Aqueous methanol is commonly used for NMR-based metabolomics because it can extract both polar and mid-polar compounds (Hendrawati *et al*., 2006; Yang *et al.,* 2006; Leiss *et al*., 2009). Attempts to design a universal extraction protocol have been made and these led to an isolation protocol in which both aqueous methanol and chloroform are used in order to target both polar and non-polar metabolites (Choi *et al*., 2004b; Choi *et al.,* 2005).

Another important factor to be considered during metabolite extraction is the pH of the medium/solvents which should be maintained from one extraction to another in order to minimize ionization of metabolites which could lead to the greatest variability especially in techniques such as NMR (Kim *et al*., 2010). To overcome such problems, buffers are normally used to maintain the pH, for instance phosphate buffers (Kim *et al*., 2006) and formic acid is also used to enhance ionization for LC-MS (Theodoridis *et al*., 2011)**.** It is therefore important that strict measures are taken into consideration during sample handling and preparation in order to minimize the problems associated with such steps during metabolic analyses which can lead to false conclusions.

2.3.3. Analytical Platforms

Different analytical techniques have been used to analyze complex mixtures and depending on the mode of separation, each technique results in different data output. Once again, due to highly diverse physico-chemical properties of metabolites (and subsequently the complex multi-dimensionality of samples) there is currently no single analytical platform that can provide complete analysis of the whole metabolome. Hence, a multi-platform approach is advantageous so as to provide maximal metabolic coverage (Dunn et al. 2012). However different techniques such as Gas Chromatography coupled to Mass Spectrometry (GC-MS) (Olivier and Loots, 2012), Liquid Chromatography coupled to Mass Spectrometry (LC-MS) (Lu *et al*., 2008), Capillary Electrophoresis coupled to Mass Spectrometry (CE-MS), Fourier Transform Infra-red (FT-IR) spectroscopy, Direct Infusion Mass Spectrometry (DI-MS) and lastly, Nuclear Magnetic Resonance (NMR) spectroscopy (Kim *et al*., 2010) are commonly used, and more comprehensive details on these and other techniques used for metabolomics has been reported

before (Dunn *et al*., 2005). As mentioned before, each technique has advantages and short-falls, and it is therefore important that the choice of instrumentation is highly in line with the expected outcomes of the study. For the scope of this thesis, Ultra-performance Liquid Chromatography coupled to High Definition Mass Spectrometry (UHPLC-HD-MS) was used, and full information on the usability of this instrument is discussed below.

2.3.3.1. Ultra-Performance Liquid Chromatography coupled to Mass Spectrometry (UHPLC-MS)

In general, mass spectrometry (MS) is regarded as a very sensitive technique to analyze metabolites (Dunn *et al*., 2005). Although complex, MS instruments are composed of two main sections namely the **(1)** ionization technology of which the following are mostly used; electron ionization (EI), chemical ionization (CI), electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), matrix-assisted laser desorption/ionization (MALDI) or fast atom/ion bombardment (FAB) and **(2)** the type of mass analyzer of which the mostly used are; quadrupole (Q), triple-quadrupole (QQQ), ion-trap (IT) or time of flight (TOF) (Villas-Boas *et al*., 2005). In order to optimize the transmission of ions to the analyzer and detector, all three MS components are maintained under vacuum. The detected ions are recorded as pairs of *m/z* and abundance values, processed, and displayed in a mass spectral format (Moco *et al*., 2007; Allwood and Goodacre, 2010). To date, various technologies ranging from ionization methods, mass analysers to detectors, have been developed for different applications with the intent to enhance detection efficiency, mass resolving power and mass accuracy of MS (Hao and March, 2001; Tang *et al*., 2006; Moco *et al*., 2007; Brucker and Rathbone, 2010).

As mentioned before, samples can be analyzed using different chromatographic techniques in conjunction with an MS detector. To date, GC and LC are the most widely used chromatographic techniques. GC-MS has been widely used for metabolite analysis and this is due to its high separation efficiency which suffices separation of very complex biological mixtures. In addition, identification of metabolites / compounds by GC-MS is relatively easier (Villas-Boas *et al*., 2005). However GC-MS has some limitations of its own which hamper its applicability in metabolomics, one of which is associated with sample preparation. GC can only analyze samples which are volatiles in nature, and those which are not volatiles need to be converted to volatiles

before analysis through a process known as derivatization. This additional treatment leads to unwanted extra steps and prolongs the sample preparation and analysis time. Derivitization is also known to result in more complex sample-handling and an increased variance in the analysis. Identification of unknown derivatized compounds can be difficult because they are chemically modified and possess / contain different structural features which are not in commercial libraries (Villas-Boas *et al*., 2005). Another disadvantage of using GC-MS associated with the afore-mentioned notion is that not all compounds reach sufficient detectable concentration in the gas phase in the GC column at a given temperature. For instance, some high molecular weight compounds are thermally decomposed before reaching a measurable vapor pressure, and it is also known that extreme polar compounds require higher temperatures (due to strong inter-molecular forces), and such higher temperatures will result in thermal decomposition of the compound. GC instruments are therefore commonly used for smaller, low polarity compounds (Figure 1.12) (Kaal and Jansen, 2008; Castello *et al*., 2009).

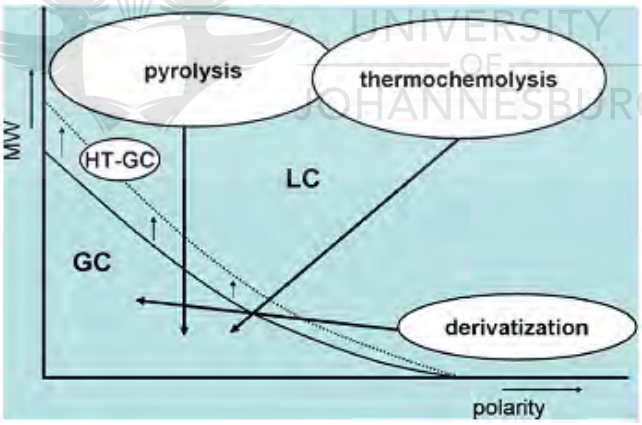


Figure 1.12: A schematic representation of the GC and LC analytical applicability ranges. From the figure it can be seen that polarity and molecular size are major determinants/factors influencing the choice of which type of instrument is sufficient for particular application (Taken from Kaal and Jansen, 2008).

With the above mentioned pitfalls associated with GC-MS, LC provides a second option for metabolomics. From this figure, it can also be noted that LC-based separation is possible across compounds of wide range of polarity and size (De Vos *et al*., 2007, Kaal and Jansen, 2008, Xiao *et al*., 2012). In recent years, LC-based methods have become the technique of choice for metabolic analyses (Grata *et al*., 2008) and have also been proven to be superior to the GC-based

counterpart. To illustrate this point, a parallel metabolomic study was conducted by GC-MS and UHPLC-MS to investigate the patho-physiology of irritable bowel syndrome. It was shown that the UHPLC-based method was three times more sensitive and detected more metabolites than its counterpart GC-based method (Kajander *et al*., 2009). LC-MS allows molecular identification and quantification of a wide spectrum of metabolites of different polarity and neutral metabolites, even when they are present at relatively low levels and in a complex matrix (Villas-Boas *et al*., 2005). LC-based techniques has been intensively used as either stand-alone techniques (Plumb *et al*., 2002; Plumb *et al*., 2005, Wagner *et al*., 2006), or in combination with other techniques such as NMR spectroscopy (Lenz *et al*., 2004, 2005a&b). There exist different types of LC-based methods and they differ with some of the instrumental components which are developing with time as to improve the detection and throughput of the technique. Such LC-based techniques include, amongst others, the conventional HPLC-MS, capillary HPLC and lastly the newly developed UHPLC-MS. The use of HPLC-MS for metabol (n)omics studies is finding many applications. Moreover, the availability of such instruments has resulted in a rapid and continuing increase in the number of publications lately (Wilson *et al*., 2005b).

In general, the use of HPLC-MS for metabolomic studies has been performed using reverse-phase packing materials of 3-5 µm particle size and in columns which range between 3.0 and 4.6 mm in diameter, and with the length between 5 and 25 cm (Wilson *et al*., 2005b). Although not common, the use of normal phase columns has also been tried for metabolomics studies. For example, some of the biological bio-fluids such as urine contain highly polar molecules that are not suitable for reversed phase chromatography. For such and other reasons, normal phase techniques is favored for the analysis of such samples and this lead to the development of the hydrophilic interaction liquid chromatography (HILIC), which result in the elution of less polar molecules first and then retention of more polar molecules. The metabolomic study by Idborg *et al.* (2005) used HILIC chromatography in the analysis of urine samples with some success. The same technique (HILIC chromatography) has also been used in conjunction with ESI-MS for the analysis of dichloroacetic acid in rat blood and tissues (Delinsky *et al*., 2005), and for the determination of 5-fluorouracil in plasma and tissues with APCI based mass spectrometry (Pisano *et al*., 2005). In the context of plant sciences, HILIC was previously shown to be an efficient method of choice during the analysis of defense-related molecules (glucosinolates) in

cruciferous vegetables. This method was shown to be superior (as it uses intact glucosinolates) over traditional reverse phase chromatography which involves an extra-tedious step involving the enzymatic (hydrolytic) removal of the sulphate group prior chromatographic analysis (Troyer *et al*., 2001).

Conventional HPLC-MS based methods have been shown to provide reasonable resolution and moderate throughput. However, developments in the techniques have seen the introduction of capillary HPLC based methods, which comes with certain improvements for metabolomic studies as compared to conventional HPLC. The capillary HPLC technique uses monolithic columns which have chromatographic resolutions approaching that of GC (Dunn, 2008) and increased signal to noise (S/N) due to more concentrated peaks and reduced ion suppression (Lu *et al*., 2008). This technique is also known to use less sample volumes than conventional HPLC. Its application for plant-based metabolomics has also yielded some positive results in the past (Delinsky *et al*., 2005; Shen *et al*., 2005; Maruska and Kornysova, 2006). Granger and associates (2005), showed that conventional and capillary HPLC columns, packed with the same chromatographic phase, yielded different results when urine from male and female rats were analyzed simultaneously by the two techniques. It was further shown in this study that the overall pattern of peak distribution provided by both techniques is significantly different when considering the pool of ions detected in both methods (Granger *et al*., 2005). In summary, capillary HPLC-based methods are hailed as superior compared to conventional HPLC methods based on several factors of which some were addressed by Lu *et al*. (2008).

Another development on the LC-based method is the introduction of UHPLC-MS. These technological advances have seen the application of LC-based methods which use sub-2 µm chromatographic particles which provide higher chromatographic resolution, narrower peak widths and higher sensitivities (Plumb and Wilson, 2004; Swartz, 2005). Due to this extremely small column particle sizes, there is an increased back-pressure which in turn affects the chromatography, and as such, these instruments are designed to handle such pressures and normally operate at up to 15 000 psi as compared to 6 000 psi which is custom for conventional HPLC (Wilson *et al*., 2005a). Different column chemistries and column dimensions have been

applied in UHPLC-MS and the application thereof in metabolomics is well documented (Wang *et al.*, 2011a).

As mentioned before, UHPLC-based methods produce superior results compared those achieved by traditional LC methods. This can be elaborated on by using an experiment conducted by Gika *et al.* (2008) as an illustrative example, whereby the authors performed global metabolite profiling of Zucker obese rats with UHPLC-MS and traditional reversed-phase LC. From their findings it was concluded that chromatography by UHPLC-MS provided better results than the conventional reversed-phase LC. This is due to high peak capacity and better peak asymmetry attained with UHPLC-MS (Gika *et al.*, 2008). In a separate but similar example, Wilson and colleagues (2005a) compared the UHPLC and HPLC based methods using similar analytical conditions. The UHPLC–MS method showed improved phenotypic classification capability and increased ability to probe differential pathway activities between strains as a result of improved analytical sensitivity and resolution (Wilson *et al.*, 2005a) above all the total ions detected by UHPLC (2.1 cm×100 mm Waters ACQUITY 1.7 µm C_{18} column) were also found to be higher than those achieved with HPLC (2.1 cm×100 mm Waters Symmetry 3.5µm C_{18} column) (Figure 1.13).

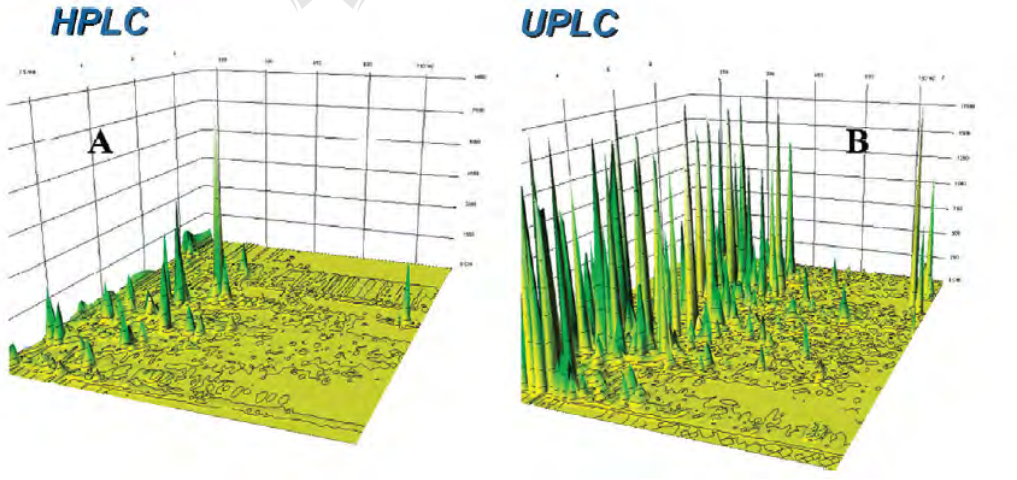


Figure 1.13: Three dimensional plots showing the retention time, *m/z*, and intensity of samples from mouse urine using (A) traditional HPLC-MS and (B) UHPLC-MS (Wilson *et al.*, 2005a).

From the above results it is quite clear that UHPLC provides superior resolution compared to traditional LC-based methods, and arguably the column chemistry and the use of small particles

stands out as one of the major improvements in this techniques (Wyndham *et al*., 2003). It is from this improved chemistry and orientation of the column materials that the running time during chromatographic separation by UHPLC is significantly shorter.

In comparison with 5 µm column particles, which are custom for traditional HPLC, it is believed that the running time is decreased (three times shorter) when 3 µm particle sizes are utilized (Novakova *et al*., 2006). The work by Plumb and associates (2005) illustrated the above point. The authors showed that using 1.5 min run times for UHPLC, the peak capacity and number of detected ions were quite similar to those achieved by 10 min HPLC run times (Plumb *et al*., 2005). It is also noteworthy that UHPLC displayed better retention time reproducibility over HPLC, which makes it an ideal technique for metabolomic analysis (Nordstrom *et al*., 2006). Some of the column chemistries used by the Waters Company for the Acquity UHPLC instruments are shown in Figure 1.14, and in the current study the BEH C18 column was used. There are more reports which are mostly concerned with the evaluation of stability and chromatographic efficacy of these types of column material and so far most of these reports consider these materials as superior both for metabolomics and other application such as in pharmaceutical industry (Wyndham *et al*., 2003; Nguyen *et al*., 2006; Novakova *et al*., 2006; Guillarme *et al*., 2007).

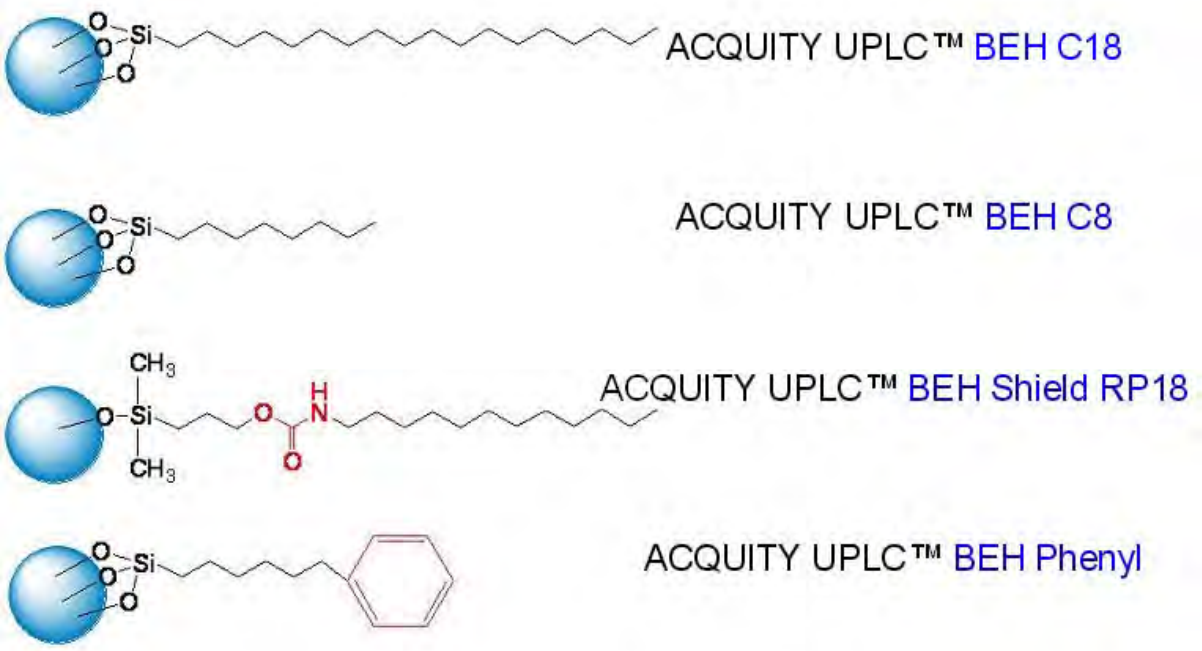


Figure 1.14: Different ethylene bridged hybrid column chemistries used in Acquity columns for UHPLC analyses.

Mass spectrometry provides significant advantages for its application in metabolomics studies (Villas-Boas *et al*., 2005; Dettmer *et al*., 2007) and as such, techniques hyphenated to MS have found significant attention for metabolic studies, with UHPLC being regarded as a golden technique for metabolomics. Furthermore, in their review on UHPLC-MS and metabolomics, Wang *et al*., (2011b), concluded that the high selectivity and low-detection limits of the MS instruments as well as the compatibility with UHPLC separation techniques and the unquestionable ability to generate quantitative data makes it an ideal tool for metabolomic applications. As previously stated, there are three main mass analyzers: TOF, quadrupole and ion trap, which are known to provide significant advantages in metabolomics. However, TOF is regarded the best amongst them because of its capability to detect all the ions simultaneously regardless of the respective masses. The other two mass analysers (quadrupole and ion trap) are superior when dealing with selected mass ranges (Dunn, 2008). Thus, TOF based MS-instruments are finding many applications in metabolomics and they have proven ideal for such applications. Hybrid MS instruments such as those comprising both quadrupole and TOF (i.e Q-TOF) have also found application in metabolomics (Plumb *et al*., 2002, 2003a&b; Williams *et al*., 2005).

In the current study such MS-instrument (Q-TOF) was also used and the diagram of its components are shown in Figure 1.15. From this schematic representation it can be seen that this instrument is a hybrid instrument capable of performing accurate mass measurements as well as tandem MS (MS/MS) experiments for structural elucidation for both precursor and product ions. From the same figure it can also be noted that there exists a hight pressure cell between the two analysers (Q and TOF) known as the collision-induced dissociation (CID) cell, which is aimed at aiding in further dissociation of selected ions from the quadrupole prior entering the TOF, thus generating more structural feature information (Dettmer *et al*., 2007; Dunn, 2008).

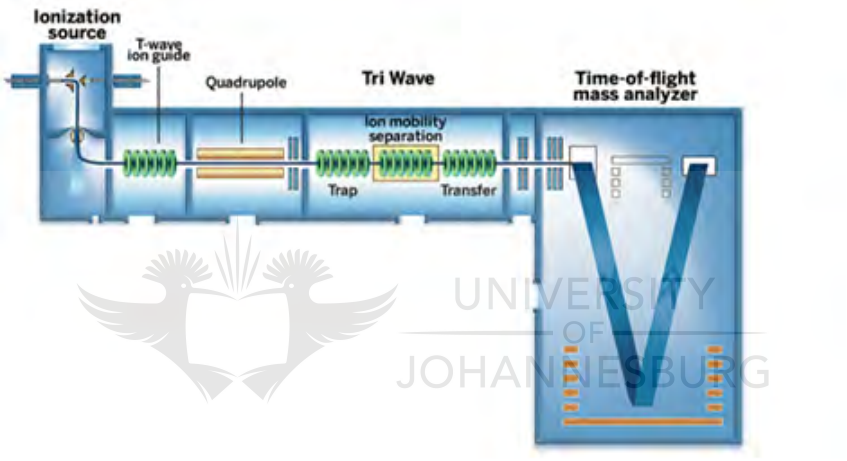


Figure 1.15: Schematic representation outlining instrumentation layout of the q-TOF-ESI-MS instrument (www.waters.com).

2.3.4. Data Handling and Analyses

Due to the large complexity and multi-dimensionality of the data obtained during metabolomic analyses, it is difficult and demanding to extract meaningful information from such data. Multivariate statistical models such as PCA (2.3.4.1) and OPLS-DA (2.3.4.2) are commonly used to reduce the dimensionality of these large data sets, providing models that are meaningfully interpretable. These models allow visualization of the data in 2- or 3-dimension (which is initially impossible in the multi-dimensional, original data) and by so doing, the underlying patterns within the data can be explained.

Before the dimensionality of the data is reduced, data pre-treatment is the initial and probably most important step. This allows transformation of the data into a format which is compatible with most of the statistical packages used to analyze the data. In the current study commercial data mining software (Markerlynx and AMIX) were used to retrieve data from UHPLC-MS chromatograms and NMR spectra respectively. These two software packages function differently. Markerlynx uses an ApexTrack peak detection technology to integrate peaks from LC-MS data: each peak on the chromatogram is given an identifier, which is mainly characterized by the area under the peak, retention time (Rt) and mass (*m/z*). Following peak detection, the peaks from different samples are aligned so that the same peaks (Rt, *m/z*), i.e. most probably the same compound, are found in the same row for all samples. During this process the retention times are allowed to differ by ± 0.2 min and the *m/z* values by ± 0.5 Thomson (Idborg *et al*., 2005).

On other hand, AMIX is used for NMR data and in this procedure, the NMR spectrum is divided into a series of small bins (buckets). The sum of intensities of signals in each bin is calculated either by relative intensities to reference areas or to the sum of total intensities (Kim *et al*., 2010). Binning can be used to both reduce the size or dimensionality of data and to accommodate small differences in the peak shift caused by pH variation, and to ensure all samples included in pattern recognition analyses are corrected for such variation (Holmes *et al*., 1994; Craig *et al*., 2006). Bin size of 0.04 ppm are normally used (Dao *et al*., 2009; Kim *et al*., 2010) and provides a good compromise between spectral resolution and positional variation of resonances.

Furthermore, scaling is another important step before data analysis. Scaling ensures that the statistical models are not excessively influenced by variables which are predominant or of high concentration as these typically have larger degrees of variance relative to variables which are less predominant (Van de Berg *et al*., 2006; Roberts, 2010). Scaling is achieved by multiplying each variable by a specific scaling weighting. There are several types of scaling procedures, however, for the scope of the thesis only two will be discussed. Before scaling, the data is mean-centered by subtracting the respective mean for each data variable. However, after mean-centering of the data, variables with large intensities would remain over-emphasized over those with less intensity. Scaling methods such as unit-variance (UV) in which each variable is divided

by its respective standard deviation is used, and this results in uniform or balanced data, giving all variables an equal weight, irrespective of respective initial intensity. Pareto scaling, on other hand, is achieved by dividing each variable by the square root of its respective standard deviation and this renders greater weight to variables which originally had larger intensity, but is not as extreme as using un-scaled data (Eriksson *et al*., 2001; Craig *et al*., 2006; Van den Berg *et al*., 2006; Loo *et al*., 2010;). In the current study, Pareto scaling was used, unless stated otherwise.

2.3.4.1. Principal Component Analysis (PCA)

PCA is a non-supervised mathematical procedure which reduces the dimensionality of data without altering the data itself. PCA is mathematically defined as an orthogonal linear transformation of possibly correlated variables into a smaller number of uncorrelated variables called principal components (PCs), where greatest variance within the data by any projection is explained on the first coordinate (called the first principal component), and the least variance is explained/projected by subsequent PCs (Jolliffe, 2002; Liu *et al*., 2010). Figure 1.16 illustrates schematically this simplified-definition of PCA: each observation is plotted in k-dimensional space, each dimension corresponding to an individual variable. The PCs must intercept the origin and have the minimal square of the distance between each observation and itself. The observations are projected onto the PCs to generate a coordinate; this value is the score, t_1, for PC 1 or t_2 for PC 2 (see Figure 1.16). In other words, PCA converts the data obtained by high-throughput instruments into a simple visual representations known as score plots which show the data as the clustering of biological samples into either similar or different groupings. Here, sample data from different biological backgrounds are clearly separated into distinct clusters and samples that cluster together can be referred as a specific "metabolic phenotype" (Fiehn *et al*., 2000).

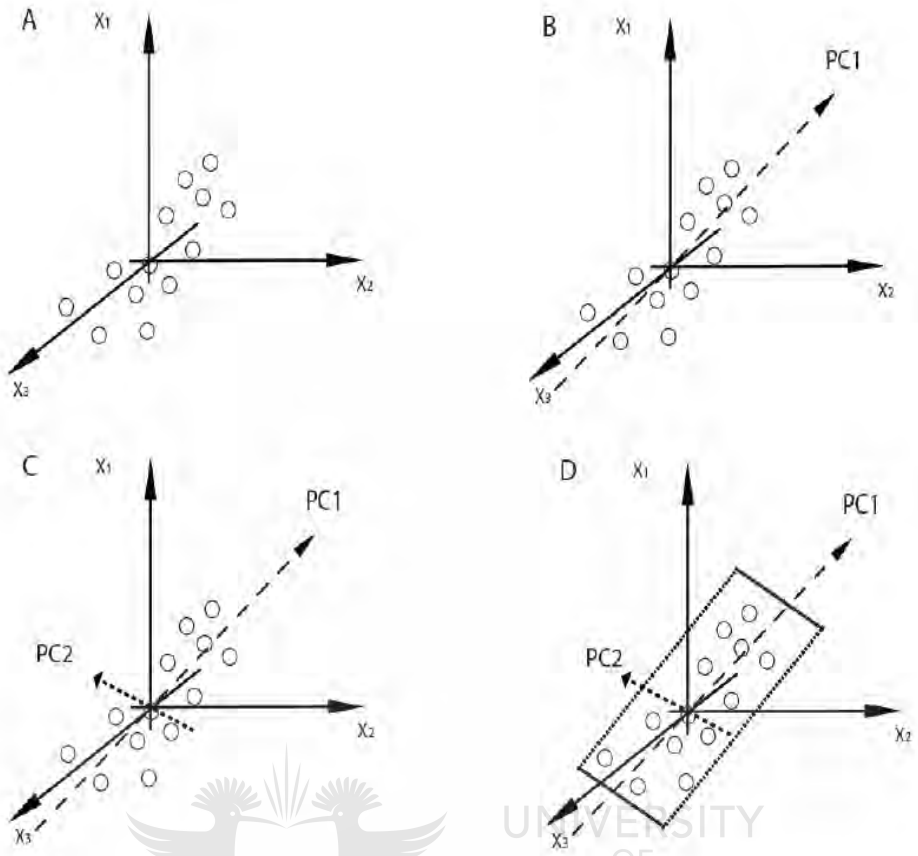


Figure 1.16: Schematic representation showing how principal component analysis is performed on metabolomics data. **A**, data is projected into three dimensional space with X_1, X_2, and X_3 being the original variables; **B**, the first principal component (PC1) is projected through the data, this is done in such a way that maximum variation with the data is explained; **C**, the second component (PC2) is also projected perpendicularly to the first one, so that the second most variation is also encompassed; **D**, a plane is generated by the two PCs in the variable space (adapted from Roberts, 2010).

2.3.4.2. Orthogonal Projections to Latent Structure Discriminant Analysis (OPLS-DA)

It is clear that PCA only evaluates global patterns (maximum variation) within the data and is not a good tool for revealing local phenomena. For the same and other reasons stated by Van der Greef and Smilde (2005), alternative techniques have been proposed. In this study a supervised model, OPLS-DA, was used to reveal underlying biological differences which are associated with a time-dependant INAP treatment as shown by the shared and unique structures (SUS) plot. OPLS-DA can be considered as an extension of the traditional PLS-DA, and it was proposed solely to handle the class orthogonal variation (hence the name) in the data which is not explained by PLS-DA (Trygg and Wold, 2002, Bylesjo *et al*., 2006, Wiklund *et al*., 2008). In detail, PLS-DA is a linear regression model which seeks to find relationships between two respective data tables (***X*** and ***Y***), where ***X*** is normally the instrument derived/measured data (i.e. GC/LC-MS or NMR data) while ***Y*** represents a binary vector which is associated with class membership. OPLS-DA separates the variation on ***X*** into two parts, one that is linearly related to ***X*** and another one that is orthogonal to ***Y***. This fragmentation on the *X* variation results in OPLS-DA comprised of two variations, one which is the ***Y***-predictive variation and explains the variation between class membership, and the other which is the ***Y***-orthogonal, which explain the variation within the class membership (Wiklund *et al*., 2008).

1.4. References

Abdulrazzak NI; Abdulrazzak N; Pollet B; Ehlting J; Larsen K; Asnaghi C; Ronseau S; Proux C; Erhardt M; Seltzer V; Renou J-P; Ullmann P; Pauly M; Lapierre C; Werck-Reichhart D (2006) A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. Plant Physiology 140: 30–48.

Almeida C; Part N; Bouhired S; Kehraus S; König JM (2011) Stachylines A-D from the sponge-derived fungus *Stachylidium sp*. Journal of Natural Products 74: 21–25.

Ackermann BL; Hale JE; Duffin KL (2006) The role of mass spectrometry in biomarker discovery and measurement. Current Drug Metabolism 7: 525–539.

Allwood JW; Goodacre R (2010) An introduction to liquid chromatography–mass spectrometry instrumentation applied in plants metabolomic analyses. Phytochemical Analysis 21: 33–47.

Alvarez ME; Nota F; Cambiagno DA (2010) Epigenetic control of plant immunity. Molecular Plant Pathology 11: 563–576.

Alvarez-Sánchez B; Priego-Capote F; Luque de Castro MD (2010) Metabolomics analysis II: Preparation of biological samples prior to detection. Trends in Analytical Chemistry 29: 120–127.

Alvarez-Venegas R; Al Abdallat A; Guo M; Alfano JR; Avramova Z (2007) Epigenetic control of a transcription factor at the cross section of two antagonistic pathways. Epigenetics 2: 106–113.

Bak S; Kahn, RA; Nielsen HL; Møller BL; Halkier BA (1998) Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Molecular Biology 36: 393–405.

Bak S; Olsen CE; Halkier BA; Møller BL (2000) Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in Dhurrin biosynthesis. Plant Physiology 123: 1437–1448.

Ball CA; Brazma A (2006) MGED standards: work in progress. Omics 10: 138–144.

Beckers GJM; Jaskiewicz M; Liu Y; Underwood WR; He YH; Zhang S; Conrath U (2009) Mitogen-activated protein kinase 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana.* Plant Cell 21: 944–953.

Beecher CWW (2004) The human metabolome. In: Harrigan GG and Goodacre R (ed.). Metabolic Profiling: It's Role in Biomarker Discovery and Gene Function Analysis, Kluwer Academic Publishers, London; 311–319.

Beets CA; Huang J-C; Madala NE; Dubery IA (2012) Activation of camalexin biosynthesis in *Arabidopsis thaliana* in response to perception of bacterial lipopolysaccharides: a gene-to-metabolite study. Planta 136: 261–272.

Bednarek P (2012) Chemical warfare or modulators of defense responses – the function of secondary metabolites in plant immunity. Current Opinion in Plant Biology: 1–8.

Bednarek P; Pis´ lewska-Bednarek M; Ver Loren van Themaat E; Maddula RK; Svatosˇ A; Schulze-Lefert P (2011) Conservation and clade-specific diversification of pathogen-inducible tryptophan and indole glucosinolate metabolism in Arabidopsis thaliana relatives. New Phytologist 192: 713–726.

Bell JD; Brown JCC; Sadler PJ (1989) NMR studies of body fluids. NMR in Biomedicine 2: 246–256.

Bennet RN; Wallsgrove RM (1994) Secondary metabolism in plant defense mechanisms. New Phytologist 127: 617–633.

Bentley R (1990) The shikimate pathway - A metabolic tree with many branches. Critical Review of Biochemistry and Molecular Biology 25: 307–338.

Berr A; McCallum EJ; Alioua A; Heintz D; Heitz T; Shen WH (2010) Arabidopsis histone methyltransferase set domain group 8 mediates induction of the jasmonate/ethylene pathway genes in plant defense response to necrotrophic fungi. Plant Physiology 154: 1403–1414.

Bezhani S; Winter C; Hershman S; Wagner JD; Kennedy JF; Kwon CS; Pfluger J; Su Y; Wagner D (2007) Unique, shared, and redundant roles for the Arabidopsis SWI/SNF chromatin remodeling ATPases brahma and splayed. Plant Cell 19: 403–416.

Bollina V; Kushalappa AC; Choo TM; Dion Y; Rioux S (2011) Identification of metabolites related to mechanisms of resistance in barley against *Fusarium graminearum*, based on mass spectrometry. Plant Molecular Biology 77: 355-370

Boyko A; Kathiria P; Zemp FJ; Yao Y; Pogribny I; Kovalchuk I (2007) Transgenerational changes in the genome stability and methylation in pathogen-infected plants (virus-induced plant genome instability). Nucleic Acids Research 35: 1714–1725.

Broadhurst DI; Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2: 171–196.

Brucker GA; Rathbone GJ (2010) Autoresonant trap mass spectrometry (ART MS) for remote sensing applications. International Journal of Mass Spectrometry 295: 133–137.

Bylesjo B; Rantalainen M; Cloarec O; Nicholson JK; Holmes E; Trygg J (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics 20: 341-351.

Castello G; Moretti P; Vezzani S (2009) Retention models for programmed gas chromatography. Journal of Chromatography A 1216: 1607–1623.

Chen C; Gonzalez FJ; Idle JR (2007) LC–MS-based metabolomics in drug metabolism. Drug Metabolism Reviews 39: 581–597.

Chinnusamy V; Zhu JK (2009) Epigenetic regulation of stress responses in plants. Current Opinion in Plant Biology 12: 1–7.

Chisholm ST; Coaker G; Day B; Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124: 803–814.

Choi HK; Choi YH; Verberne M; Lefeber AW; Erkelens C; Verpoorte R (2004a) Metabolic fingerprinting of wild type and transgenic tobacco plants by 1H NMR and multivariate analysis technique. Phytochemistry 65: 857–864.

Choi YH; Kim HK; Hazekamp A; Erkelens C; Lefeber AWM; Verpoorte R (2004b) Metabolomic differentiation of *Cannabis sativa* cultivars using 1H NMR spectroscopy and principal component analysis. Journal of Natural Products 67: 953–957.

Choi YH; Sertic S; Kim HK; Wilson EG; Michopoulos F; Lefeber AWM; Erkelens C; Prat Kricun SD; Verpoorte R (2005) Classification of Ilex species based on metabolomic fingerprinting using NMR and multivariate data analysis. Journal of Agricultural and Food Chemistry 53: 1237–1245.

Choi YH; Spronsen JV; Dai Y; Verberne M; Hollmann F; Arends IWCE; Witkamp G-Jan; Verpoorte R (2011) Are natural deep eutectic solvents the missing link in understanding cellular metabolism and physiology. Plant Physiology 156: 1701–1705.

Choi YH: Kim HK; Linthorst HJM; Hollander JG; Lefeber AWM; Erkelens C; Nuzillard JM; Verpoorte R (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. Journal of Natural Products 69: 742–748.

Conrath U; Pieterse CMJ; Mauch-Mani B (2002) Priming in plant pathogen interactions. Trends in Plant Science 7: 210–216.

Conrath U; Beckers GJM; Flors V; Garcia-Augustin P; Jakab G; Mauch F; Newman MA; Pieterse CM; Poinssot B; Pozo MJ; Pugin A; Schaffrath U; Ton J; Wendehenne D; Zimmerli L; Mauch-Mani B (2006) Priming getting ready for battle. Molecular Plant Microbe Interaction 19: 1062–1071.

Craig A; Cloarec O; Holmes E; Nicholson JK; Lindon JC (2006) Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical Chemistry 78: 2262–2267.

Dao TTH; Chacon R; Kyong H; Erkelens C; Lefeber AWM; Linthorst HJM; Hae Y; Verpoorte R (2009) Plant physiology and biochemistry effect of benzothiadiazole on the metabolome of *Arabidopsis thaliana*. Plant Physiology and Biochemistry 47: 146–152.

Dawson GW; Hallahan DL; Mudd A; Patel MM; Pickett JA; Wadhams LJ; Wallsgrove RM (1989) Secondary plant metabolites as targets for genetic modification of crop plants for pest resistance. Pesticide Science 27: 191–201.

De Ascensao; ARFDC; Dubery, IA. (2003) Soluble and wall-bound phenolics and phenolic polymers in *Musa acuminata* roots exposed to elicitors from *Fusarium oxysporum* f.sp. cubense. Phytochemistry 63: 679-686.

Dempsey DA; Klessing DF (2012) SOS-too many signals for systemic acquired resistance. Trends in Plant Science 17: 538–545.

Delinsky AD; Delinsky DC; Muralidhara S; Fisher JW; Bruckner JV; Bartlett MG (2005) Analysis of dichloroacetic acid in rat blood and tissues by hydrophilic interaction liquid chromatography with tandem spectrometry. Rapid Communication in Mass Spectrometry 19: 1075–1083.

Dettmer K; Aronov PA; Hammock BD (2007) mass spectrometry-based metabolomics. Mass Spectrometry Reviews 26: 51–78.

De Vos RC; Moco S; Lommen A; Keurentjes JJB; Bino RJ; Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2: 778–791.

Dieterle F; Schlotterbeck GT: Ross A; Niederhauser U; Senn H (2006) Application of metabonomics in a compound ranking study in early drug development revealing drug-induced excretion of choline into urine. Chemical Research in Toxicology 19: 1175–1181.

Dixon RA (2001) Natural products and plant disease resistance. Nature 411: 843–847

Dixon RA; Paiva NL (1995) Stress induced phenylpropanoid metabolism. Plant Cell 7: 1085–1097.

Dixon RA; Lamb CJ; Masoud S; Sewalt VJ; Paiva NL (1996) Metabolic engineering: prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses – a review. Gene 179: 61–71.

Dohleman FG; Long SP (2009) More productive than maize in the mid-west: how does Miscanthus do it ? Plant Physiology 150: 2104–2115.

Dudareva N; and Pichersky E (2000) Biochemical and molecular genetic aspects of floral scent. Plant Physiology 122: 627–633.

Dunn WB (2008) Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. Physical Biology 5: 011001.

Dunn WB; Bailey NJC; Johnson HE (2005) Measuring the metabolome: current analytical Technologies. Analyst 130: 606–625.

Dunn WB; Erban A; Weber RJM; Creek DJ; Brown M; Breitling R, Hankemeier T, Goodacre R, Neumann S; Kopka J; Viant MR (2012a) Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics *DOI* 10.1007/s11306-012-0434-4.

Eriksson L; Johansson E; Kettaneh-Wold N; Trygg J; Wikström C; Wold S (2006) Multi- and megavariate data analysis; Principles and applications. Volume 1, First edition, Umetric Academy: Umea: Sweden.

Faoro F; Maffi D; Cantu D; Iriti M (2008) Chemical-induced resistance against powdery mildew in barley: the effects of chitosan and benzothiadiazole. Biocontrol 53: 387–401.

Fava F; Lovegrove J A; Gitau R; Jackson K G; Tuohy KM (2006) The gut microbiota and lipid metabolism: implications for human health and coronary heart disease Current Medical Chemistry 13: 3005–3021.

Fernie AR; Trethewey RN;Krotzky AJ; Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nature Reviews Molecular Cell Biololgy 5: 763–769.

Fiehn O; Kopka J; Dormann P; Altmann; T; Trethewey;R.N; Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nature Biotechnology 18: 1157–1161.

Fiehn O; Robertson D; Griffin J; van der Werf M; Nikolau B; Morrison N; Sumner LW; Goodacre R; Hardy NW; Taylor C; Fostel J; Kristal B; Kaddurah-Daouk R; Mendes P; van Ommen B; Lindon JC; Sansone SA (2007) The metabolomics standards initiative (MSI). Metabolomics 3: 175–178.

Friedrich L; Lawton K; Ruess W; Masner P; Specker N; Rella MG; Meier B; Dincher S; Staub T; Uknes S; Me´traux JP; Kessmann H; Ryals J (1996) A benzothiadiazole derivative induces systemic acquired resistance in tobacco. The Plant Journal 10: 61–70.

Frost CJ; Mescher MC; Carlson JE; De Moraes CM (2008) Plant defense priming against herbivores: Getting ready for a different battle. Plant Physiology 146: 818-824.

Games DE; van der Greef J; Nijssen LM; Maarse H; Ten Noever de Brauw MC (1984) Analysis of pepper and capsicum oleoresins by high performance liquid chromatography- mass spectrometry and field desorption mass spectrometry. Journal of Chromatographic Science 294: 269–279.

Garcia-brugger A; Lamotte O; Vandelle E; Bourque S; Lecourieux D; Poinssot B; Wendehenne D; Pugin A (2006) Early signaling events Induced by elicitors of plant defenses. Molecular Plant Microbe Intercations 19: 711–724.

Gates SC; Sweeley CHC (1978) Quantitative metabolic profiling based on gas chromatography. Clinical Chemistry 24: 1663–1673.

Gatz C; Lenk I (1998) Promoters that respond to chemical inducers. Trends in Plant Science 3: 352–358.

German JB; Watkins SM; Fay LB (2005) Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health. Journal of the American Dietetic Association 105: 1425–1432.

Gika HG; Theodoridis GA; Wilson ID (2008) Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. Journal of separation science 31: 1598-1608.

Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology 43: 205–227.

Goellner K; Conrath U (2008) Priming: It's all the world to induced disease resistance. European Journal of Plant Pathology 121: 233–242.

Gomez-Ariza J; Campo S; Rufat M; Estopa M; Messeguer J; Segundo BS; Coca M (2007) Sucrose-mediated priming of plant defense responses and broad spectrum disease resistance by overexpression of the maize pathogenesis related proteins in rice plants. Molecular Plant Microbe Interactions 20: 363–369.

Goodacre R (2007) Metabolomics of a superorganism. Journal of Nutrition 137: 259–266.

Gorlach J; Volrath S; Knauf-Beiter G; Hengy G; Beckhove U; Kogel KH; Oostendorp M; Staub T; Ward E; Kessmann H; Ryals J (1996) Benzothiadiazole, a novel class of inducers of systemic acquired resistance, activates gene expression and disease resistance in wheat. Plant Cell 8: 629–643.

Granger J; Plumb R; Castro-Perez J; Wilson ID (2005) Metabonomic studies comparing capillary and conventional HPLC-oa-TOF MS for the analysis of urine from zucker obese rats. Chromatographia 61: 375-380.

Grata E; Boccard J; Guillarme D; Glauser G; Carrupt P-alain; Farmer EE; Wolfender J-L; Rudaz S (2008) UHPLC–TOF-MS for plant metabolomics: A sequential approach for wound marker analysis in *Arabidopsis thaliana*. Journal of Chromatography B 871: 261–270.

Griffin JL; Nicholls AW; Keun HC; Mortishire-Smith RJ; Nicholson JK; Kuehn T (2002) Metabolic profiling of rodent biological fluids via 1H NMR spectroscopy using a 1 mm microlitre probe. Analyst 127: 582–584.

Guillarme D; Nguyen DT; Rudaz S; Veuthey J-luc (2007) Recent developments in liquid chromatography—Impact on qualitative and quantitative performance. Journal of Chromatography A 1149: 20-29.

Hagemeier J; Schneider B; Oldham NJ; Hahlbrock K (2001) Accumulation of soluble and wall-bound indolic metabolites in *Arabidopsis thaliana* leaves infected with Virulent or avirulent *Pseudomonas syringae* pathovar tomato strains. Proceedings of National Academy of Science 98: 753–758.

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytologist 169: 453–468.

Hammerschmidt R (1999) Phytoalexins: what have we learned after 60 years? Annual Review of Phytopathology 37: 285–306.

Hammond-Kosack KE, Jones JDG (1996) Resistance gene-dependant plant defense responses. Plant Cell 8: 1773-1791.

Hanhineva K (2008) Metabolic engineering of phenolic biosynthesis pathway and metabolite profiling of strawberry (*Fragaria* × *ananassa*). Kuopio University Publications C. Natural and Environmental Sciences 231: 89.

Hao C; March RE (2001) A survey of recent research activity in quadrupole ion trap mass spectrometry. International Journal of Mass Spectrometry 212: 337–357.

Haslam E (1993) Shikimic acid: metabolosm and metabolites. John Wiley, Chichester; New York, 157–168.

Harborne JB (1999) The comparative biochemistry of phytoalexin induction in plants. Biochemical Systematics and Ecology 27: 335–367.

Hartmann T (2007) From waste products to ecochemicals: 50 years research on plant secondary metabolism. Phytochemistry 68: 2831–2846.

Hendrawati O; Yao Q; Kim HK; Linthorst HJM; Erkelens C; Lefeber, AWM; Choi, YH; Verpoorte R (2006) Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. Plant Science 170: 1118–1124.

Hermann KM (1995) This shikimate pathway: early steps in the biosynthesis of aromatic compounds. Plant Cell 7: 907–919.

Herrmann KM; Weaver LM (1999) The shikimate pathway. Annual Review Plant Physiology and Plant Molecular Bioliogy 50: 473–503.

Holmes E; Foxall PJ; Nicholson JK; Neild GH; Brown SM (1995) Automatic data reduction and pattern recognition methods for analysis of 1H nuclear magnetic resonance spectra of human urine from normal and pathological states. Analytical Biochemistry 220: 284–296.

Holmes E; Loo RL; Cloarec O; Coen M; Tang H; Maibaum E; Bruce S; Chan Q; Elliott P; Stamler J; Wilson ID; Lindon JC; Nicholson JK (2007) Detection of urinary drug metabolite (Xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. Analytical Chemistry 79: 2629–2640.

Huang HC; Schuurink R; Denny TP; Atkinson MM; Baker CJ; Yucel I; Hutcheson SW; Collmer A (1988) Molecular cloning of a *Pseudomonas syringae pv. syringae* gene cluster that enables *Pseudomonas fluorescens* to elicit the hypersensitive response in tobacco plants. Journal of Bacteriology 170: 4748–4756.

Idborg H; Zamani L; Edlund PO; Schuppe-Koistinen I; Jacobsson SP (2005) Metabolic fingerprinting of rat urine by LC/MS Part 1: Analysis by hydrophilic interaction liquid chromatography–electrospray ionization mass spectrometry. Journal of Chromatography B 828: 9–13.

Ioannidis JPA (2005) Why most published research findings are false?. PLoS Medicine 2:696–701.

Iriti M; Faoro F (2003) Benzothiadiazole (BTH) induces cell-death independent resistance in *Phaseolus vulgaris* against Uromyces appendiculatus. Journal of Phytopathology 151: 171–180.

Iriti M; Faoro F (2009) Chemical diversity and defense metabolism: how plants cope with pathogens and ozone pollution. International Journal of Molecular Sciences 10: 3371–3399.

Jahangir M; Abdel-Farid IB; Kim HK; Choi YH; Verpoorte R (2009) Healthy and unhealthy plants: the effect of stress on the metabolism of Brassicaceae. Environmental and Experimental Botany 67: 23–33.

Jellum E (1977) Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry with special reference to organic acids. Journal of Chromatography 143: 427–462.

Jellum E (2001) Chromatography, mass spectrometry and electrophoresis for diagnosis of human disease, particulary metabolic disorders. In Chromatography—A Century of Discovery 1900–2000, Gehrke Ch, Wixom R, Bayer E (eds). Elseviers Science, Netherlands; 270–277.

Jensen RA (1986) Tyrosine and phenylalanine biosynthesis: relationship between alternative pathaways, regulation and subcellular. In Recent advances in phytochemistry: The shikimic acid pathway (Conn E.E, Ed) Plenum Press, New York, 57–81.

Jolliffe IT (2002) Principal Component Analysis. 2nd edition, Springer, New York.

Jones JD; Dangl JL (2006) The plant immune system. Nature 444: 323–329.

Jørgensen K; Rasmussen AV; Morant M; Bak S; Møller BL (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Current Opinion in Plant Biology 8: 280–291

Jung HW; Tschaplinski TJ; Wang L; Glazebrook J; Greenberg1 JT (2009) Priming in systemic plant immunity. Science 324: 89–91.

Kaal E; Janssen HG (2008) Extending the molecular application range of gas chromatography. Journal of Chromatography A 1184: 43–60.

Kathiria P; Sidler C; Golubov A; Kalischuk; Kawchuk LM; Kovalchuk I (2010) Tobacco mosaic virus infection results in an increase in recombination frequency and resistance to viral, bacterial, and fungal pathogens in the progeny of infected tobacco plants. Plant Physiology 153: 1859–1870.

Kajander K; Myllyluoma E; Kyro npalo S; Rasmus- sen M; Sipponen P; Mattila I; Seppanen-Laakso T; Vapaatalo H; Oresic M; Korpela R; World J (2009) Elevated pro-inflammatory and lipotoxic mucosal lipids characterise irritable bowel syndrome. World Journal Gastroenterology 15: 6068–6142.

Kell DB (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Discovery Today 11: 1085–1092.

Kell DB; Oliver SG (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26: 99–105.

Khoo SHG; Al-RubeaiM (2007) Metabolomics as a complementary tool in cell culture. Biotechnology and Applied. Biochemistry 47: 71–84.

Kim HK; Choi YH; Verpoorte R (2006) Metabolomic analysis of *Catharanthus roseus* using NMR and principal component analysis. In Biotechnology in Agriculture and Forestry. Plant Metabolomics (eds. Saito, K., Dixon, R.A. & Willmitzer, L.) 57: 261–276.

Kim HK; Choi YH; Verpoorte R (2010) NMR-based metabolomic analysis of plants. Nature protocols 5: 536-549.

Kim HK; Verpoorte R (2010) Sample preparation for plant metabolomics. Phytochemical Analysis 21: 4–13.

Kinross JM; Holmes E; Darzi AW; Nicholson JK (2011) Metabolic phenotyping for monitoring surgical patients. Lancet 377:1817–1819.

Koch BM; Sibbesen O; Halkier BA; Svendsen I; Møller BL (1995) The primary sequence of cytochrome P450tyr, the multifunctional N- hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. Archives of Biochemistry and Biophysics 323: 177–186.

Koek MM; Muilwijk B; van der Werf MJ; Hankemeier T (2006) Microbial metabolomics with gas chromatography/ mass spectrometry. Analytical Chemistry 78: 1272–1281.

Kohler A; Schwindling S; Conrath U (2002) Benzothiadiazole-induced priming for potentiated responses to pathogen infection, wounding and infiltration of water into leaves requires the NPR1/NIM1 gene in Arabidopsis. Plant Physiology 128: 1046–1056.

Kopka J; Fernie A; Weckwerth W; Gibon Y; Stitt M (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biology 5: 109.

Kruger NJ; Huddleston JE; Le Lay P; Brown ND; Ratcliffe RG (2007) Network flux analysis: impact of ^{13}C-substrates on metabolism in *Arabidopsis thaliana* cell suspension cultures. Phytochemistry 68: 2176–2188.

Kruger NJ; Troncoso-Ponce AT; Ratcliffe RG (2008) 1H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. Nature Protocols 3: 1001–1012.

Kuc J (1982) Induced immunity to plant disease. Biosciences 32: 854-860

La Camera S; Gouzerh G; Dhondt S; Hoffmann L; Frittig B; Legrand M; Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunology Reviews 4: 198:267-284.

Laiakis EC; Morris GAJ; Fornace AJ Jr; Howie SRC (2010) Metabolomic Analysis in Severe Childhood Pneumonia in The Gambia, West Africa: Findings from a Pilot Study. PLoS ONE 5: e12655.

Lawton KA; Friedrich L; Hunt M; Weymann K; Delaney T; Kessmann H; Staub T; Ryals T (1996) Benzothiadiazole induces disease resistance in Arabidopsis by activation of the systemic acquired resistance signal transduction pathway. The Plant Journal 10: 71–82.

Leiss KA; Choi YH; Abdel-Farid IB; Verpoorte R; Klinkhamer PGL (2009) NMR Metabolomics of thrips (*Frankliniella occidentalis*) resistance in senecio hybrids. Journal of Chemical Ecology 35: 219–229.

Lenz EM; Bright J; Knight R; Wilson ID; Major H (2005a) A metabonomic investigation of the biochemical effects of mercuric chloride in the rat using 1H NMR and HPLC-TOF/MS: time

dependent changes in the urinary profile of endogenous metabolites as a result of nephrotoxicity. Analyst 129: 535–541.

Lenz EM; Bright J; Knight R; Wilson ID; Major H (2004) Cyclosporin A-induced changes in endogenous metabolites in rat urine: a metabonomic investigation using high field 1H NMR spectroscopy, HPLC-TOF/MS and chemometrics. Journal of Pharmaceutical and Biomedical. Analysis 35: 599–608.

Lenz EM; Bright J; Knight R; Westwood FR; Davies D; Major H;Wilson ID (2005b) Metabonomics with 1H NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. Biomarkers 10: 173–187.

Lindon JC; Holmes E; Nicholson JK; (2003) So what's the deal with metabonomics? Analytical Chemistry 75: 384A–391A.

Lindon JC; Holmes; Nicholson JK (2006) Metabonomics techniques and applications to pharmaceutical research and development. Pharmaceutical Research 23: 1075–1088.

Lindon JC; Nicholson JK (2008) Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. Trends in Analytical Chemistry 27:194–204.

Lin CY; Viant MR; Tjeerdema RS (2006) Metabolomics: methodologies and applications in the environmental sciences. Journal of Pesticide Science 31:245–251.

Liu NQ; Cao M; Frédérich M; Choi Y-H; Verpoorte R; Van der Kooy F (2010) Metabolomic investigation of the ethnopharmacological use of *Artemisia afra* with NMR spectroscopy and multivariate data analysis. Journal of Ethnopharmacology 128: 230–235

Luna E; Bruce TJA; Roberts MR; Flors V; Ton J (2011) Next generation systemic acquired resistance. Plant Physiology 158: 844–853.

Lu X; Zhao X; Bai C; Zhao C; Lu G; Xu G (2008) LC–MS-based metabonomics analysis. Journal of Chromatography B 866: 64–76.

Maffi D; Iriti M; Pigni M; Vannini C; Faoro F (2011) *Uromyces appendiculatus* infection in BTH-treated bean plants: ultrastructural details of a lost fight. Mycopathologia 171: 209–221.

Maltese F; van der Kooy F; Verpoorte R (2009) Solvents derived artifacts in natural product chemistry. Natural Product Communications 4: 447–454.

Maldonado AM; Doerner P; Dixon RA; Lamb CJ; and Cameron RK (2002) A putative lipid transfer protein involved in systemic resistance signaling in Arabidopsis. Nature 419: 399–403.

March-Diaz R; Garcia-Dominguez M; Lozano-Juste J; Leon J; Florencio FJ; Reyes JC (2008) Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in Arabidopsis. The Plant Journal 53: 475–487.

Maruska A; Kornysova O (2006) Application of monolithic (continuous bed) chromatographic columns in phytochemical analysis. Journal of Chromatography A 1112: 319–3130.

Mashego MR; Rumbold K; DeMey M; Vandamme E; Soetaert W; Heijnen JJ (2007) Microbial metabolomics: past, present and future methodologies. Biotechnology Letters 29: 1–16.

McDowell JM; Dangl JL (2000) Signal transduction in the plant immune response. Trends in Biochemical Sciences 25: 79-82.

Meuzelaar HLC; Kistemaker PG (1973) A technique for fast and reproducible fingerprinting of bacteria by pyrolysis mass spectrometry. Analytical Chemistry 45: 587–590.

Mishina TE; Zeier J (2007) Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis. The Plant Journal 50: 500–513.

Moco S; Vervoort J (2007) Metabolomics technologies and metabolite identification. Trends in Analytical Chemistry 26: 855–866.

Moller BL (2010) Dynamic metabolons. Science 330: 1328-1329.

Nanda T; Das M; Tripathy K; Ravi Teja, Y (2011) Metabolomics: The Future of Systems Biology. Journal of Computer Sciences and Systems Biology. Doi:10.4172/jcsb.R1-003

Nguyen DT; Guillarme D; Rudaz S; Veuthey J-luc (2006) Chromatographic behaviour and comparison of column packed with sub-2 µm stationary phases in liquid chromatography. Chromatographia 1128: 105–113.

Nicholson JK; Lindon JC; Holmes E (1998) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29: 1181–1189

Nicholson JK; Wilson ID (1989) High resolution proton magnetic resonance spectroscopy of biological fluids. Progress NMR Spectroscopy 21: 449–501.

Nielsen KA; Tattersall DB; Jones PR (2008) Metabolon formation in dhurrin biosynthesis. Molecular Engineering 69: 88–98.

Nordstrom A; O'Maille G; Qin C; Siuzdak G (2006) Non-linear data alignment for UHPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. Analytical Chemistry 78: 3289–3384.

Novakova L; Matysov L; Solich P (2006) Advantages of application of UHPLC in pharmaceutical analysis. Talanta 68: 908–918.

Odjakova M; Hadjiivanova C (2001) The complexity of pathogen defense in plants. Bulgarian Journal of Plant Physiology 27: 101-109.

Olivier I; Loots DT (2012) A metabolomics approach to characterize and identify various Mycobacterium species. Journal of Microbiological Methods 88: 419–426.

Oliver SG; Winson MK; Kell DB; Baganz F (1998) Systematic functional analysis of the yeast genome. Trends in Biotechnology 16: 373–378.

Osbourn AE (1996) Preformed antimicrobial compounds and plant defense against fungal attack. Plant Cell 8: 1821–1831.

Padidam M (2003) Chemically regulated gene expression in plants. Current Opinion in Plant Biology 6: 169–177.

Pastor V; Luna E; Ton J; Flors V (2012) Primed plants do not forget. Environmental and Experimental Botany. doi:10.1016/j.envexpbot.2012.02.013.

Petersen M; Abdullah Y; Eberle D; Gehlen K; Hücherig S; Janiak V; Kim KH; Sander M; Wetzel C; Wolters S (2009) Evolution of rosmarinic acid biosynthesis. Phytochemistry 70: 1663–1679.

Pichersky E; Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends in plant science 5: 439-445.

Pieterse CMJ; Leon-Reyes A; Van der Ent S; Van Wees SCM (2009) Networking by small-molecule hormones in plant immunity. Nature Chemistry and Biology 5: 308–316.

Pisano R; Breda M; Grassi S; James CA (2005) Hydrophilic interaction liquid chromatography-APCI-mass spectrometry determination of 5-fluorouracil in plasma and tissues. Pharmaceutical and Biomedical Analysis 38: 738–745.

Plumb RS; Granger JH; Stumpf CL; Johnson KA; Smith BW; Gaulitz S; Wilson ID; Castro-Perez J (2005) A rapid screening approach to metabonomics using UHPLC and oa-TOF mass spectrometry : application to age, gender diurnal variation in normal/zucker obese rats and black, white and nude mice. Analyst 130: 844–849.

Plumb RS; Wilson ID (2004) High throughput and high sensitivity LC/MS–OA–TOF and UHPLC/TOF–MS for the identification of biomarkers of toxicity and disease using a metabonomics approach. Abstracts of Paper of American Chemical Society 228: U164–U.

Plumb R; Granger J; Stumpf C; Wilson ID; Evans JA; Lenz EM (2003a) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOF-MS): Detection of strain, diurnal and gender differences. Analyst 128: 819–823.

Plumb RS; Johnson KA; Rainville P; Smith BW; Wilson ID; Castro-perez JM; Nicholson JK. (2006) UHPLC / MS^E; a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Communications in Mass Spectrometry 20: 1989–1994.

Plumb RS; Stumpf CL; Gorenstein MV; Castro-Perez JM; Dear GJ; Anthony M; Sweatman BC; Connor SC; Haselden JN (2002) Metabonomics: The use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. Rapid Communication in Mass Spectrometry 16: 1991–1996.

Plumb RS; Stumpf CL; Granger JH; Castro-Perez J; Haselden JN; Dear GJ (2003b) Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. Rapid Communication in Mass Spectrometry 17: 2632–2638.

Politzer IA; Dowty BJ; Laseter JL (1976) Use of gas chromatography and mass spectrometry to analyze underivatised volatile human or animal constituents of clinical interest. Clinical Chemistry 22: 1775–1788.

Pyo JS; Ju HK; Park JH; Kwon SW (2008) Determination of volatile biomarkers for apoptosis and necrosis by solid-phase microextraction-gas chromatography/mass spectrometry: A pharmacometabolomic approach to cisplatin's cytotoxicity to human lung cancer cell lines. Journal of Chromatography B 876: 170–174.

Queiroz O (1974) Circadian-rhythms and metabolic patterns. Annual Review Plant Physiology and Plant Molecular Biology 25: 115–134.

Ramamoorthy V; Viswanathan R; Raguchander T; Prakasam V; Samiyappan R (2001) Introduction to systemic resistance by plant growth promoting rhizobacteria in crop plants against pests and diseases. Crop Protection 20: 1–11.

Ransohoff DF (2005) Opinion-bias as a threat to the validity of cancer molecular-marker research. Nature Reviews Cancer 5: 142–149.

Roberts LD (2010) Defining the metabolic effect of peroxisome proliferator-activated receptor δ activation. Doctoral thesis; University of Cambridge.

Robertson DG (2005) Metabonomics in Toxicology: A Review Toxicological Science 85: 809–822.

Ryals JA; Neuenschwander UH; Willits MG; Molina A; Steiner H; Hunt MD (1996) Systemic acquired resistance. Plant Cell 8: 1809-1819.

Ryan D; Robards K (2006) Metabolomics: the greatest omics of them all? Analytical Chemistry 78: 7954–7958.

Sano H (2010) Inheritance of acquired traits in plants: reinstatement of Lamarck. Plant Signal Behavior 5: 346–348.

Schilmiller AL; Stout J; Weng JK; Humphreys J; Rueger MO; Chapple CCS (2009) Mutations in the CINNAMATE 4-HYDROXYLASE gene impact metabolism, growth and development in Arabidopsis. The Plant Journal 60: 771–782.

Schmidt J; Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. Phytochemistry 39: 737–749.

Schoch G; Morant M; Abdulrazzak N; Asnaghi C; Goepfert S; Petersen M; Ullmann P; Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environmental Chemistry Letters 4: 127–136.

Seger C; Sturm S (2007) Analytical aspects of plant metabolite profiling platforms: Current standings and future aims. Journal of Proteome Research 6: 480–497.

Shen YF; Zhang R; Moore RJ; Kim J; Metz TO; Hixson KK; Zhao R; Livesay EA; Udseth HR; Smith RD (2005) Automated 20 kpsi RPLC–MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics. Analytical Chemistry 77: 3090–100.

Slaughter A; Daniel X; Flors V; Luna E; Hohn B; Mauch-mani B (2012) Descendants of primed *Arabidopsis* plants exhibit resistance to biotic stress. Plant Physiology 158: 835–843.

Spoel SH; Dong XN (2008) Making sense of hormone crosstalk during plant immune responses. Cell Host and Microbe 3: 348-351.

Swartz ME (2005) UHPLC (TM): an introduction and review. Journal of Liquid Chromatography and Related Technology 28: 1253–1263.

Tan JW; Bednarek P; Liu HK; Schneider B; Svatos A; Hahlbrock K (2004) Universally occurring phenylpropanoid and species-specific indolic metabolites in infected and uninfected *Arabidopsis thaliana* roots and leaves. Phytochemistry 65: 691–699.

Tang X; Bruce JE; Hill HH (2006) Characterizing electrospray ionization using atmospheric pressure ion mobility spectrometry. Analytical Chemistry 78: 7751–7760.

Tao Y; Xie Z; Chen W; Glazebrook J; Chang H-S; Han B; Zhu T; Zou G; Katagiri F (2003) Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae.* Plant Cell 15: 317-330.

Tas AC; van der Greef J (1995) Mass spectrometric profiling and pattern recognition. Mass Spectrometry Reviews 13: 155–181.

Taylor CF; Hermjakob H; Julian RK; Garavelli JS; Aebersold R; Apweiler R (2006) The work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI) Omics 10: 145–151.

Theodoridis G; Gika HG; Wilson ID (2011) mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. Mass Spectrometry Reviews 10: 884-906.

Thilmony R; Underwood W; He SY (2006) Genome-wide transcriptional analysis of the *Arabidopsis thaliana* interaction with the plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 and the human pathogen *Escherichia coli* O157:H7. The Plant Journal 46: 34-53.

Thomma B P H J; Eggermont K; Penninckx I A M A; Mauch-Mani B; Vogelsang R; Cammue BPA; Broe-kaert W F (1998) Separate jasmonate-dependent and salicy- late-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. Proceedings of the National Academy of Science USA 95:15107–15111.

Thomma BPHJ; Nelissen I; Eggermont K; Broekaert WF (1999) Deficiency in phytoalexin production causes enhanced susceptibility of *Arabidopsis thaliana* to the fungus Alternaria brassicicola. The Plant Journal 9: 163–171.

t'Kindt R; Morreel K; Deforce D; Boerjan W; Van Bocxlaer J (2009) Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment. Journal of Chromatography B 877: 3572–3580.

Ton J; Jakab G; Toquin V; Flors V; Iavicoli A; Maeder MN; Metraux JP; Mauch-Mani B (2005) Dissecting the β- aminobutyric acid-induced priming phenomenon in Arabidopsis. Plant Cell 17:987–999.

Troyer JK, Stephenson KK; Fahey JW (2001) Analysis of glucosinolates from broccoli and other cruciferous vegetables by hydrophilic interaction liquid chromatography. Journal of Chromatography A 919: 299–304.

Trygg J; Wold S (2002) Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics 16: 119-128.

Van Deemter JJ; Zuiderweg J; Klingengerg A (1956) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Journal Chemical Engineering Science 5: 271–289.

Van den Berg RA; Hoefsloot HC; Westerhuis JA Smilde AK; Van Der Werf MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC genomics *7*: 1-15.

Van Etten HD; Mansfield JW; Bailey JA; Farmer EE (1994) Two classes of plant antibiotics: phytoalexins versus "phytoanticipins". Plant Cell 9: 1191–1192.

Van der Ent S; Van Wees SCM; Pieterse CMJ (2009) Jasmonate signaling in plant interactions with resistance-inducing beneficial microbes. Phytochemistry 70: 1581-1588.

Van der Greef J; Martin S; Juhasz P; Adourian A; Plasterer T; Verheij ER; McBurney RN (2007) The art and practice of systems biology in medicine: mapping patterns of relationships. Journal of Proteome Research 6: 1540–1559.

Van der Greef J; Smilde AK (2005) Symbiosis of chemometrics and metabolomics: past, present, and future. Journal of Chemometrics 19: 376–386.

Van der Greef J; Stroobant P; van der Heijden R (2004) The role of analytical sciences medical systems biology Current Opinion in Chemical Biology 8: 559–565.

Van der Greef J; Tas AC; Bouwman J; Ten Noever de Brauw MC; Schreurs WHP (1983) Evaluation of field desorption and fast atom bombardment mass spectrometric profiles by pattern recognition techniques. Analytical Chemica Acta 150: 45–52.

Van Loon LC (2000) Systemic induced resistance. In: A. Slusarenko, R.S.S. Fraser, L.C. van Loon (Eds), Mechanisms of resistance to plant diseases. Kluwer Academic Publishers, Netherlands, pp. 521–574.

Van Loon LC; Bakker PAHM; Pieterse CMJ (1998) Systemic resistance induced by rhizosphere bacteria. Annual Review in Phytopathology 36: 453–483.

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes and Developments 20: 759–771.

Verpoorte R; Alfermann AW (2000) secondary metabolism in metabolic engineering of plant secondary metabolism, Kluwer Academic Publshers, printed in Netherlands, 1–29.

Viant MR (2007) Metabolomics of aquatic organisms: the new 'omics' on the block. Marine Ecology Progress Series 332: 301–6.

Vicedo B; Flors V; De la; Oleyva M; Finiti I; Kravchuk Z; Real MD; García-Agustín P; González-Bosch C (2009) Hexanoic acid-induced resistance against *Botrytis cinerea* in tomato plants. Molecular Plant-Microbe Interactions 22:1455–1465.

Villas-Boas SG; Mas S; Åkesson M; Smedsgaard J; Nielsen J (2005) Mass spectrometry in metabolome analysis. Mass Spectrometry Reviews 24: 613–646.

Vogt T (2010) Phenylpropanoid biosynthesis. Molecular Plant 3: 2-20.

Wagner S; Scholz K; Donegan M; Burton L; Wingate J; Volkel W (2006) Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. Analytical Chemistry 78: 1296-1305.

Walters DR; Paterson L; Walsh DJ; Havis ND (2009) Priming for plant defense in barley provides benefits only under high disease pressure. Physiological and Molecular Plant Pathology 73: 95–100.

Wang X; Sun H; Zhang A; Sun W; Wang P; Wang Z (2011a) Potential role of metabolomics apporoaches in the area of traditional Chinese medicine: as pillars of the bridge between Chinese and Western medicine. Journal of Pharmaceutical and Biomedical Analysis 55: 859–868.

Wang X; Wang H; Zhang A; Lu X; Sun H; Dong H; Wang P (2012) Metabolomics study on the toxicity of aconite root and Its processed products using ultra-performance liquid-chromatography / electrospray-Ionization synaptic high-definition mass spectrometry coupled with pattern recognition approach and ingenuity pathways analysis. Journal of Proteome Research 11: 1284–1301.

Wang X; Zhang HAS; Wang P; Han Y (2011b) Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. Journal of Separation Science 34: 3451–3459.

Weckwerth W; Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation Drug Discovery Today 10: 1551–1558.

Wendehenne D; Durner J; Chen Z; Klessig F (1998) Benzothiadiazole, an inducer of plant defenses, inhibits catalase and ascorbate peroxidase. Phytochemistry 47: 651–657.

Winkel BSJ (2004) Metabolic channeling in plants. Annual Reviews of Plant Biology 55:85–107.

Wiklund S; Johansson E; Sjo L; Shockcor JP; Gottfries J; Moritz T; Trygg J (2008) Visualization of GC / TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical Chemistry 80: 115–122.

Williams RE; Major H; Lock EA; Lenz EM; Wilson ID (2005) D-Serine- induced nephrotoxicity: A HPLC-TOF/MS-based metabonomics approach. Toxicology 207: 179–190.

Wilson ID; Nicholson JK; Castro-Perez J; Granger JH; Johnson KA; Smith BW; Plumb RS (2005a) High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. Journal of Proteome Research 4: 591–598.

Wilson ID; Plumb R; Granger J; Major H;Williams R; Lenz EA (2005b) HPLC-MS-based methods for the study of metabonomics. Journal of Chromatography B 817: 67–76.

Wink M (1997) Special nitrogen metabolism. In: Plant Biochemsitry (Dey P.M and Harborne J.B, Eds) Academic press, San Diego, 239–486.

Wolski EA; Henriquez MA; Adam LR; Badawi M; Andreu AB; El Hadrami A; Daayf F (2010) Induction of defense genes and secondary metabolites in saskatoons (Amelanchier alnifolia Nutt.) in response to *Entomosporium mespili* using jasmonic acid and *Canada milkvetch* extracts. Environmental and Experimental Botany 68: 273–282.

Wyndham KD; Gara JEO; Walter TH; Glose KH; Lawrence NL; Alden BA; Izzo GS; Hudalla CJ; Iraneta PC (2003) Characterization and evaluation of C 18 HPLC stationary phases based on ethyl-bridged hybrid organic / inorganic particles. Analytical Chemistry 75: 6781–6788.

Xiao JF; Zhou1 B; Ressom HW (2012) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. Trends in Analytical Chemistry 32: 1–14.

Yang SY; Kim HK; Lefber AWM; Erkelens C; Angelova N; Choi YH; Verpoorte R (2006) Application of two dimensional nuclear magnetic resonance spectroscopy to quality control of ginseng commercial products. Planta Medicine 72: 364–369.

Yu K; Sheng G; Sheng J; Chen Y; Xu W; Liu X; Cao H; Qu H; Cheng Y; Li L (2007) A metabonomic investigation on the biochemical perturbation in liver failure patients caused by Hepatitis b virus. Journal of Proteome Research 6: 2413–2419.

Zhang SJ; Yang X; Sun MW; Sun F; Deng S; Dong HS (2009) Riboflavin-induced priming for pathogen defense in *Arabidopsis thaliana*. Journal of Integrative Plant Biology 51: 167–174.

Zimmerli L; Metraux JP; Mauch-Mani B (2001) β-aminobutyric acid-induced protection of Arabidopsis against the necrotrophic fungus *Botrytis cinerea*. Plant Physiology 126: 517–523.

Zimmerli L; Jakab C; Metraux JP; Mauch-Mani B (2000) Potentiation of pathogen-specific defense mechanisms in *Arabidopsis* by β-aminobutyric acid. Proceedings of the National Academy of Science USA 97: 12920–12925.

Zwiegelaar M; Dubery IA (2006) Early activation of cell wall strengthening-related gene transcription in cotton by a *Verticillium dahliae* elicitor. South African Journal of Botany 72: 467-472.

Chapter 3: The short and long of it: Shorter chromatographic analysis suffice for sample classification during UHPLC-MS-based metabolic fingerprinting.

3.1. Abstract

Ultra high-performance liquid chromatography hyphenated to mass spectrometry (UHPLC-MS) technologies has been widely applied in metabolomics, and the high resolution and peak capacity thereof are only some of the key aspects that are exploited in such and related fields. In the current study, we investigated if low resolution chromatography, with the aid of multivariate data analyses, could be sufficient for a metabolic fingerprinting study that aims at discriminating between samples of different biological status or origin. UHPLC-MS data from chemically-treated *Arabidopsis thaliana* plants were used and chromatograms with different gradient lengths compared. MarkerLynx^{TM} technology was employed for data mining, followed by principal component analysis (PCA) and orthogonal projections to latent structure discriminant analysis (OPLS-DA) as multivariate statistical interpretations. The results showed that, despite the congestion in low resolution chromatograms (of 5 and 10 min), samples could be classified based on the respective biological background in a similar manner as when using chromatograms with better resolution (of 20 and 40 min). This paper thus underlines that, in a metabolic fingerprinting study, low resolution chromatography together with multivariate data analyses suffice for biological classification of samples. The results also suggest that, depending on the initial objective of the undertaken study, optimization in chromatographic resolution prior to full scale metabolomics studies is mandatory.

Keywords

Ultra high-performance liquid chromatography - mass spectrometry, Data mining, Metabolic fingerprinting, Metabolomics, Multivariate data analysis.

3.2. Introduction

Metabolomics can be viewed as a post-genomic scientific approach or discipline that aims at the comprehensive identification and quantitation of all cellular metabolites within a biological system. Metabolomics, in comparison to other "omics" technologies, deals with the dynamic metabolome of an organism [1, 2, 3, 4, 5]. Metabolites are the end products of gene expression and define the biochemical phenotype of a cell or tissue under defined conditions. Metabolite patterns can thus provide a holistic signature of the physiological state of an organism under study as well as insightful knowledge of specific biochemical processes [6,7,8,9,10]

In this paper, we focus on metabolic fingerprinting, which is a powerful tool for exploring systematic metabolic changes and potential biomarkers in a cell/organism under certain physiological conditions. As a non-targeted methodology, the initial intention with metabolic fingerprinting is not to identify the observed metabolites, but to detect differences between biological groups under investigation by comparing patterns, signatures, or "fingerprints" of metabolites that undergo dynamic changes in response to biotic or abiotic stresses, host-pathogen interactions and genetic - or environmental alterations. Metabolic fingerprinting thus aims at sample classification by rapid, global metabolite analysis [11,12]. Most of the analytical platforms used in metabolic fingerprinting studies include IR and Raman spectroscopy-, NMR spectroscopy- and MS-based technologies. Metabolic fingerprinting is increasingly becoming a valuable tool for rapid sample classification; e.g. disease diagnostics, monitoring whole-cell biotransformations, plant taxonomy, plant-based medicines, etc. [5,13,14].

In this study, data were acquired using a high-definition MS (Waters Synapt G1). The mass spectral data provides a pattern that is most often compound specific, thus enabling the extraction of metabolic patterns or signatures in samples, and ultimately compound identification and/or structural elucidation. Such information from MS analysis provides biological knowledge of the system under investigation; for instance, the classification or separation of biological samples based on differential metabolic signatures/fingerprints. In metabolomic studies, MS is often preceded by GC or LC. A better chromatographic separation can enhance the quality of MS analysis and subsequent compound identification by reducing the complexity of the mass spectra and the matrix effect [15]. However, the question in MS-based metabolic fingerprinting (aiming

mainly at sample classification by rapid, global metabolic analysis) is to know the extent to which chromatographic separation needs to be optimised and the chromatographic run time that is sufficient.

An ultra high-performance liquid chromatography (UHPLC) step can be used prior to MS data acquisition. UHPLC offers numerous advantages over traditional HPLC, of which the most important includes higher peak capacity at the same analysis time and the ability to allow shorter chromatographic runs with similar (and sometimes even better) separation as those attained by traditional HPLC [16,17,18]. The underlying principles of the UHPLC technology are fundamentally governed by the van Deemter equation, which is an empirical formula that describes the relationship between linear velocity and plate height/column efficiency [19]. According to this equation, as the particle size decreases to less than 2 μm, there is a significant gain in efficiency, which does not diminish at increased flow rates or linear velocities [20,21,22].

Metabolic fingerprinting studies, like all other metabolomic approaches, generate high-dimensional and complex data sets which are difficult to analyse and interpret by visual inspection or any traditional univariate statistical approaches. Mathematical modelling approaches involving multivariate data analysis (MVDA) methods are therefore utilised to extract meaningful information from these large empirical data sets [22,23,24].

In this study, the principal component analysis (PCA) – an unsupervised multivariate linear model – was used for data analysis. PCA provides a means of identifying patterns in data (especially of high dimension), and expressing the data in such a way as to highlight the similarities and differences. The other main advantage of PCA is that once the patterns in data are found, it can be compressed by reducing the number of dimensions without much loss of information. In other words, PCA attempts to explain as much variation in as few components as possible [25,26,27,28,29]. Another MVDA model which was used is orthogonal projection to latent structures discriminant analysis (OPLS-DA). The latter is a modification of the PLS-DA (projection to latent structures-discriminant analysis) method, with an integral orthogonal signal correction filter. The power of this regression model lies in its ability to separate modelling of response-related (predictive) and response-orthogonal variations in data. As such, the OPLS-DA model is a suitable tool to extract information on changes in the molecular composition of samples. The OPLS-DA scatter plots, S-plot and the shared-and-unique-structures (SUS)-plot

enable the extraction of statistically and biochemically potentially significant metabolites and the identification of shared and unique structures in the samples, respectively [27,30,31]. This study seeks to shed some light on these aspects by investigating the effect of short gradient/low resolution in comparison to longer gradient/high resolution chromatographic separation during metabolic fingerprinting of *Arabidopsis* plants treated with isonitrosoacetophenone (INAP), a structural analogue of citaldoxime, an anti-oxidant/anti-fungal stress metabolite from citrus [32].

3.3. Experimental

3.3.1. Chemical treatment and metabolite extraction.

Thirty (30) day old *Arabidopsis thalina* (ecotype Columbia) plants were used and foliar sprayed with 1 mM of INAP (Sigma-Aldrich, MO, USA) in 10 mM $MgCl_2$, while control plants were only sprayed with 10 mM $MgCl_2$. Plants were allowed to incubate for a period of 18 h at 25ºC prior to metabolite extraction. Leaves (2 g) were homogenised with an ultraturrax homogeniser in 20 mL 100% methanol (Romil, Cambridge, UK), and metabolites were extracted with the aid of heating at 60ºC for 10 min and sonification for 20 min. Homogenates were then centrifuged at 10 000 x *g* for 10 min, and the resulting supernatants transferred to a round bottom flask prior to reducing the volume to approximately 2 mL with a rotary evaporator. The resulting volume was further dried to completeness using a speed vacuum centrifuge (R.C 10.09; Jouan, France) operating with constant heating at 50ºC. The residues were re-dissolved in 400 µL 50% (*v/v*) methanol in water and filtered through a 0.22 µm filter using a 1 mL sterile syringe. Unless stated elsewhere, all reagents and solvents were of UHPLC grade.

3.3.2. UHPLC-HD-ESI-MS analysis

Five (5) µL methanol extract was analyzed on a Waters UHPLC-high definition MS instrument equipped with an Acquity CSH C18 column (150 mm × 2.1 mm with a particle size of 1.7 µm) (Waters Corporation, Milford, MA, USA). The composition of mobile phase A consisted of 0.1% formic acid in deionised water and mobile phase B consisted of 0.1% formic acid in

methanol. The time of the chromatographic separations were 5, 10, 20 and 40 min; with gradient conditions (methanol: 5-95%, at constant flow rate of 0.4 mL min^{-1}). Full chromatographic gradient conditions for each chromatographic analysis time are indicated in **Table 1**.

Table 1: Chromatographic gradient conditions for each chromatographic analysis time; 5, 10, 20 and 40 min. (A = 0.1% formic acid in deionised water and B = 0.1% formic acid in methanol).

Time (min)				Flow rate (mL min^{-1})	% A	% B
Initial	Initial	Initial	Initial	0.400	95.0	5.0
0.10	1.00	1.00	1.00	0.400	95.0	5.0
2.00	6.00	16.00	36.00	0.400	5.0	95.0
3.00	7.00	17.00	37.00	0.400	5.0	95.0
3.50	8.00	18.00	38.00	0.400	95.0	5.0
5.00	10.00	20.00	40.00	0.400	95.0	5.0

Data was acquired using detection with both PAD (100-500 nm) and MS. MS was used in both positive and negative electrospray ionisation modes, and for the scope of this paper only the positive ionisation data was used. For MS detection, the optimal experimental conditions were as follows: capillary voltage of 2.5 kV, sample cone voltage of 17 V, multichannel plate detector voltage of 1750 V, source temperature of 120°C, desolvation temperature of 400°C, cone gas flow of 50 L h^{-1} and desolvation gas flow of 450 L h^{-1}. The mass spectrometric full scan data were acquired from 100-1000 Da with a scan time of 0.1 sec, interscan delay of 0.02 sec. Data was centroided and mass spectra corrected in real time by an external reference standard consisting of leucine-enkephalin (5 pg mL^{-1}) using a lockmass sprayer interface and a lockmass flow rate of 0.2 mL min^{-1}.

3.3.3. Data analysis

MassLynx XS™ software version 4.1 (Waters Corporation, Milford, MA, USA) with an advanced statistical programme for multivariate data analysis (MarkerLynx™) was used to analyse the UHPLC-ESI-MS data. For multivariate data analysis, ESI positive raw data was extracted and analysed with MarkerLynx™ software (Waters Corporation, Milford, MA, USA). MarkerLynx™ parameters were set to analyse between 0.5-4 min, 2.5-8.5 min, 3-19 min and 5-

38 min for the 5, 10, 20 and 40 min runs respectively. The mass range was set to 100-1000 Da, the mass tolerance to 0.01 Da, the t_R window to 0.2 min and the mass window to 0.02 Da. Isotopic peaks were excluded from the analysis. The signal believed to be of residual INAP was also removed for the rest of the analysis. The dataset obtained from MarkerLynx™ processing was exported to the SIMCA-P software version 13.0 (Umetrics, Umea, Sweden) programme for PCA analysis. Unless stated otherwise, PCA and OPLS-DA models were centered and then Pareto scaled using SIMCA-P software. Using the loading S-plot from OPLS-DA, biomarkers (metabolites) of which the levels were highly perturbed by the treatment, thus with a correlation coefficient (P(corr)) of ≥ 0.8 and covariance coefficient (p1) ≥ 0.05, were selected and the respective *m/z* values compared using a Venn diagram [33]. The latter was performed in order to evaluate whether different lengths of chromatographic separation resulted in unique or different metabolite distribution patterns.

3.4. Results and Discussion

Visual inspection of the base peak intensity (BPI) chromatograms of different chromatographic run times (5, 10, 20 and 40 min) showed differential chromatographic resolution (Figure 1). Metabolites co-elute in a relatively small chromatographic time window (t_R of 1.5-3.5 min) when a 5 min run is considered while there is enhanced separation of detected compounds across the longer time window of 15-38 min in the 40 min runs. The three dimensional (3D) BPI chromatograms (Electronic supplementary Figure S1) also clearly indicate that when shorter (5-10 min) chromatography run times are utilized, there is peak co-elution as compared to longer (20-40 min) chromatographic runs. Both 3D and 2D BPI chromatograms show that resolution increases with increasing chromatographic run times as it was expected.

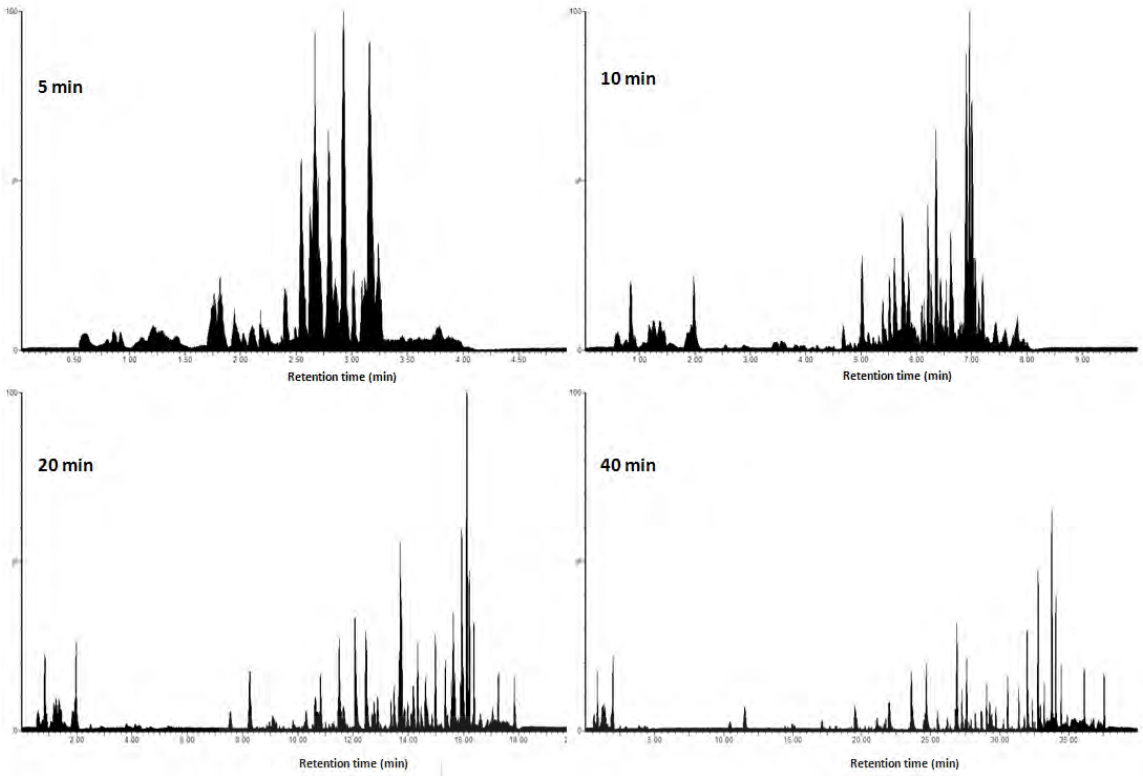


Figure 1: Representative base peak intensity (BPI) UHPLC-ESI-MS chromatograms of methanol extracts from 18 h INAP-treated *Arabidopsis thaliana* plants analyzed under different chromatographic run times.

The data were further analysed with PCA and OPLS-DA methods. PCA modelling enables the reduction of data dimensionality, thus providing interpretable visualisation of the original multi-dimensional data. The PCA scores plots show the clustering of biological samples into either similar or different groupings [16,25]. Here, the PCA models for the 5, 10 20 and 40 min UHPLC-MS data sets were respectively computed and the PCA scores plots are represented in Figure 2. From these plots, a clear separation between samples originating from the different biological backgrounds (untreated and treated) can be seen in all data sets (5, 10, 20 and 40 min runs), and the differential clustering of the samples indicates the existence of different metabolic signatures (fingerprints) in cells, indicating that INAP treatment leads to metabolic changes in *A. thaliana* plants.

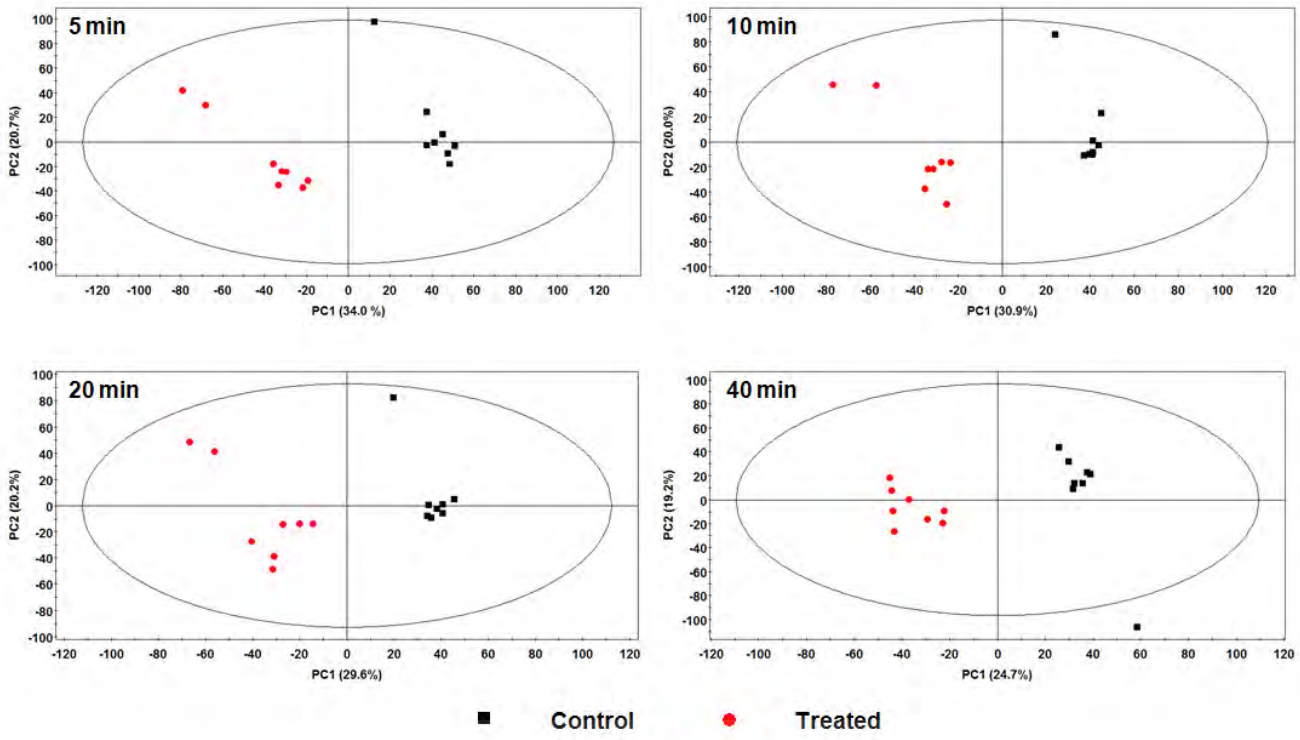


Figure 2: 2D PCA score plots showing the different clustering and separation patterns between extracts from treated and control *Arabidopsis thaliana* plants, generated by UHPLC-MS using different chromatographic running times of 5, 10, 20 and 40 min.

However, from these PCA scores plots, no significant differences in the clustering of samples (treated and untreated) was observed when data from 5, 10, 20 and 40 min chromatographic runs are compared. The only apparent difference is the tighter clustering within the treated group in 40 min samples (**Figure 2**). The PCA modelling thus suggests that despite the differential and improved chromatographic resolution obtained in longer chromatography runs (**Figure 1**), the samples obtained from INAP-treated cells were clearly separated from the samples obtained from non-treated cells in all cases (**Figure 2**).

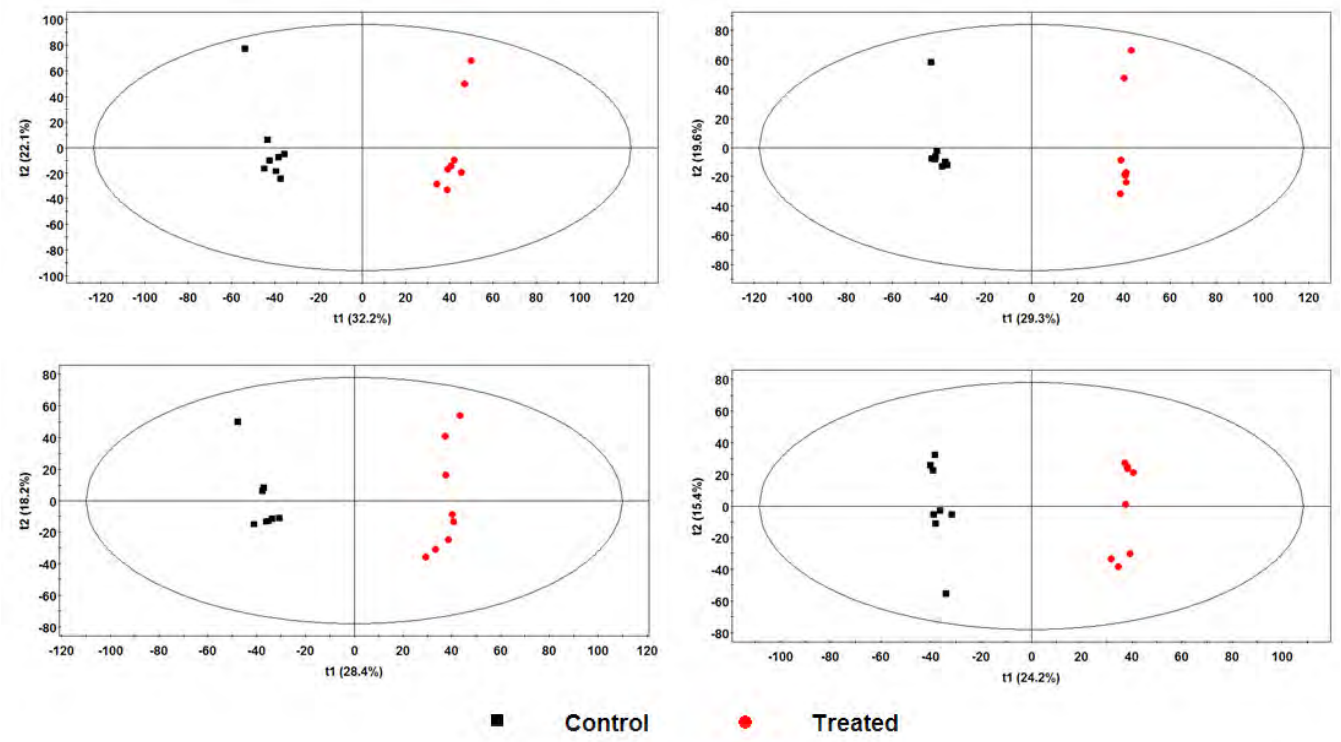


Figure 3: Representative 2D OPLS-DA score plots (**A** and **B**) and the respective loading S-plots (**C** and **D**) showing differential separation and the most influential metabolites between treated and control samples using the data from 5 min (A and C) and 40 min (B and D) chromatographic analysis times. The circled areas from the shaded boxes highlight two $[M+H]^+$ ions of which the levels are strongly affected by INAP treatment, thus contributing to the clustering/separation of groups as seen in A and B. These two ions (*m/z* 639.406 and 683.433) are the same in both 5 and 40 min runs.

To further characterise the results obtained by PCA, a supervised model, OPLS-DA, was used. Both OPLS-DA scores plots of 5 and 40 min chromatography showed separation and clustering of treated and non-treated samples (**Figure 3A** and **B**). To highlight *m/z* variables (potential bio-marker metabolites) that are responsible for sample clustering, OPLS-DA loading S-plots of both 5 and 40 min chromatography were computed (**Figure 3C** and **D**). Loadings S-plots uses the covariance - and correlation coefficients algorithms to highlight metabolites of which the levels are affected as a result of the biological perturbation in question [30]. Metabolites which are dispersed in the upper and lower outer regions of the "S"-like distribution shape represents those that are affiliated with treated and untreated samples, respectively. In the current study only up-

regulated metabolites (those with correlation coefficient, P(corr), of ≥ 0.8 and covariance coefficient, (p1) of ≥ 0.05, (thus less spurious variables), were used.

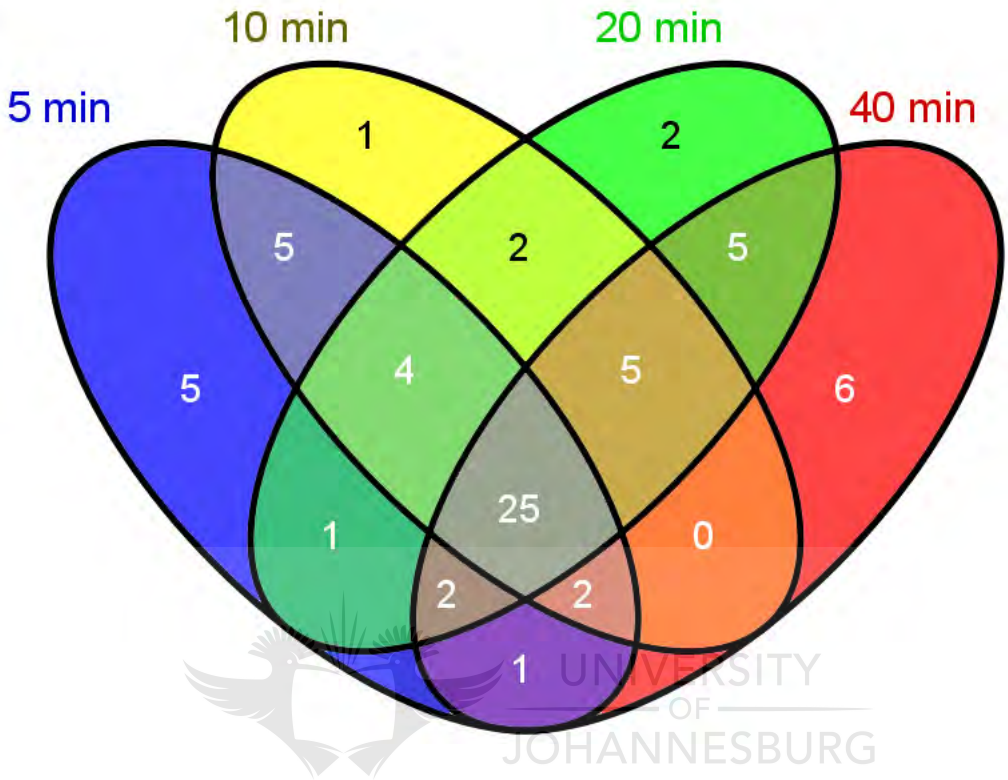


Figure 4: Venn diagram highlighting the number of unique and shared most influential biomarkers identified by UHPLC-MS that are linked to the response of *Arabidopsis thaliana* plants towards INAP, from the data generated using different chromatographic analysis times as indicated.

A close inspection of the S-plots results evidenced that in both cases of 5 and 40 min chromatography, the *m/z* ions (associated with the INAP treatment) were mostly the same (**Figure 3C** and **D,** expanded regions of most affected metabolites). This was further visualised using a Venn diagram (**Figure 4**), which allowed singling out the unique and shared *m/z* ions from data generated using differing times of chromatography separation. From this figure it is evident that varying lengths of chromatography resulted in largely similar metabolite patterns as most of the metabolites are found in the "shared" rather than unique areas. Here, 25 *m/z* ions (potential biomarkers) are common to all four datasets, thus showing high similarities in the metabolite distribution patterns. Comparing, for instance, 5 and 40 min chromatography runs, 30 *m/z* ions are shared versus the 15 or 16 *m/z* ions that are unique to either the 5 or 40 min

chromatographic runs. This indicates that, despite the 8-fold increase in the chromatographic run time, the metabolic signatures/fingerprints that were uncovered from both short and long chromatography, are 67% the same. This observation becomes more pronounced when the exclusion bar/limit is tightened by increasing the p1 value to ≥ 0.07. In the latter case most, if not all, biomarkers on the 5, 10 and 20 min runs were shared (results not shown). Thus, the separation of different biological groups (treated and untreated) on the PCA and OPLS-DA score plots of different chromatographic run times was due to the same biomarkers, as illustrated in fig.3.

Many metabolic fingerprinting or - footprinting studies have relied mostly on NMR-based approaches [34,35]. Few attempts have been carried out with LC-MS-based metabolic fingerprinting. Some of the GC-MS-based metabolic fingerprint studies used long chromatographic runs (e.g. 30 min) [12]. This study demonstrates that shorter chromatographic run times (5 min) may be sufficient for a metabolic fingerprint study that aims mainly at sample classification based on sample metabolic signatures. This observation was also seen on an separate study, conducted on tobacco cells suspensions, where the PCA results obtained from 5 min chromatographic runs were comparable with those achieved by 30 min runs (results not shown).

3.5. Conclusion

The results from this study proved that, where the only question is whether samples are different or similar (based on the metabolic signature), a shorter chromatographic analysis (coupled to a high-definition mass spectrometer) is more practical, feasible and sufficient to provide the necessary biological information. The study also demonstrated that UHPLC is a powerful technique in studying metabolomics, owing to its practical robustness and ability to produce highly complex data matrices using short analysis times. MVDA models suffice (i) the visualization of different clustering and separation of samples based on metabolite distribution patterns due to biological variations and, (ii) allows comprehensive deciphering of underlying variation due to chromatography-related factors such as analysis time.

From the current study, it is evident that a 5 min chromatographic analysis time is enough to distinguish Arabidopsis samples based on the metabolite content, using an UHPLC-MS

analytical platform. It must, however, be stressed that a generic method was used and applied over various analysis run times with no chromatographic optimisation performed as the run times changed. Such optimisation might include, amongst others, column length; in essence for such shorter runs, a shorter column would be more feasible. Another factor will be temperature that plays a significant role in chromatographic separation. Due to the small particle size utilised in the UHPLC columns, the flow rate can also be optimised in combination with the other factors mentioned above. Although a uniform set of parameters were used in the current study, excellent results were obtained. Combining a chromatography system capable high resolution with a high definition MS instrument formed the basis of a good analytical system that produced complex, but information-rich data. Specialized software such as MarkerLynx^{TM} which is capable of fully utilising the raw data by applying Apex Track^{TM} technology, contributed to downstream data analyses and - processing. Taking all of the above into consideration, it is evident that high throughput sample classification based on metabolic profiles is feasible with short chromatographic run times of 5 min.

3.6. References

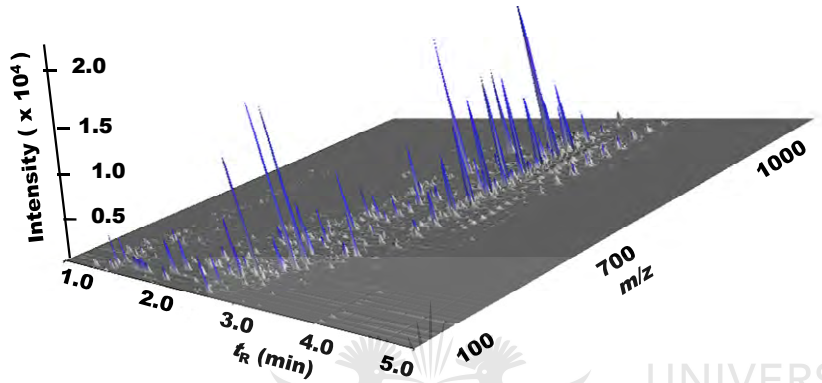
1. Hall RD (2005) New Phytol 169: 453-468
2. Oksman-Caldentey K-M, Saito K (2005) Curr Opin Biotechnol 16: 174-179
3. Verpoorte R, Choi YH, Mustafa NR, Kim HK (2008) Phytochem Rev 7: 525-539
4. Lu X, Zhao X, Bai C, Zhao C, Lu G, Xu G (2008) J Chromatogr B 866: 64-76
5. Winder CL, Dunn WB, Goodacre R (2011) Trends Microbiol 19: 315-322
6. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Nat Biotechnol 18: 1157-1161
7. Sumner LW, Mendes P, Dixon RA (2003) Phytochemistry 62: 817-836
8. Dunn WB, Ellis DI (2005) Trends Anal Chem 24 285-294
9. Fukusaki E, Kobayashi A (2005) J Biosci Bioengineer 100: 347-354.
10. Idborg H, Zamani L, Edlund P-O, Schuppe-Koistinen I, Jacobsson SP (2005) J Chromatogr B 828: 9-13
11. García-Pérez I, Vallejo M, García A, Legido-Quigley C, Barbas C (2008) J Chromatogr A 1204: 130-139
12. Zeng M, Liang Y, Li H, Wang M, Wang B, Chen X, Zhou N, Cao D, Wu J (2010) J Pharmaceut Biomed Anal 52: 265-272
13. Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R (2007) Pharmacogenomics 8: 1243-1266
14. Samra T, Sharma S, Pawar M (2011) J Clin Monitor Comp 25: 49-150
15. Xiao JF, Zhou B, Ressom HW (2012) Trends Anal Chem 32: 1-14.
16. Wilson ID, Nicholson JK, Castro-Perez J, Granger JH, Johnson KA, Smith BW, Plumb RS (2005) J Proteome Res 4: 591-598
17. Plumb RS, Granger JH, Stumpf CL, Johnson KA, Smith BW, Gaulitz S, Wilson ID, Castro-Perez J (2005) Analyst 130: 844-849.
18. Gika HG, Theodoridis G, Extance J, Edge AM, Wilson ID (2008) J Chromatogr B 871: 279-287
19. Shaaban H, Górecki T (2011) Chromatographia 74: 9-17
20. Knox JH, Scott HP (1983) J Chromatogr A 282: 297-313

21. Usher KM, Simmons CR, Dorsey JD (2008) J Chromatogr A 1200: 122-128
22. Fiehn O (2002) Plant Mol Biol 48 155-171.
23. Van den Berg RA, Rubingh CM, Westerhuis JA, van der Werf MJ, Smilde AK (2009) Anal Chim Acta 651: 173-181
24. Jansen JJ, S. Smit, H.C.J. Hoefsloot, A.K. Smilde (2010) Phytochem Anal 21: 48-60.
25. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Trends Biotechnol 22: 245-252
26. Van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ (2006) BMC Genom 7: 1-15
27. Trygg J, Holmes E, Lundstedt T (2007) J Proteome Res 6: 469-479
28. Yamamoto H, Yamaji H, Abe Y, Harada K, Waluyo D, Fukusaki E, Kondo A, Ohno H, Fukuda H (2009) Chemometr Intell Lab Syst 98: 136-142
29. Fonville JM, Bylesjö M, Coen M, Nicholson JK, Holmes E, Lindon JC, Rantalainen M (2011) Anal Chim Acta 705: 72-80
30. Wiklund S, Johansson E, Sjöström L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J (2008) Anal Chem 80: 115-122.
31. Sieber M, Wagner S, Rached E, Amberg A, Mally A, Dekant W (2009) Chem Res Toxicol 22: 1221-1231
32. Dubery IA, Louw AE, van Heerden FR (1999) Phytochemistry 50: 983-989
33. http://bioinfogp.cnb.csic.es/tools/venny/index.html accessed on 20-06-2012
34. Aliferis KA, Jabaji S (2010) Metabolomics 6: 96-108
35. Safer S, Cicek SS, Pieri V, Schwaiger S, Schneider P, Wissemann V, Stuppner H (2011) Phytochemistry 72: 1379-1389

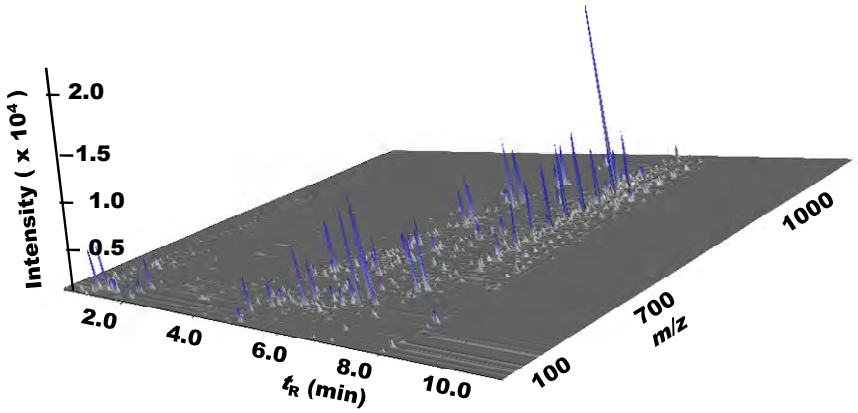
3.7. Supplementary files

Figure S1: Representative 3D UHPLC-MS BIP chromatograms generated using different chromatographic analysis times of 5, 10, 20 and 40 min.

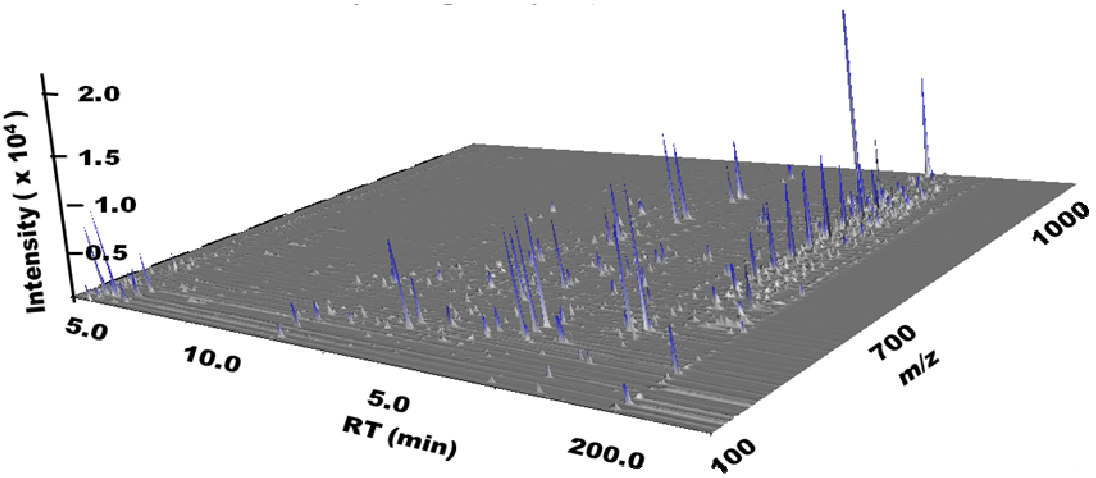
5 min



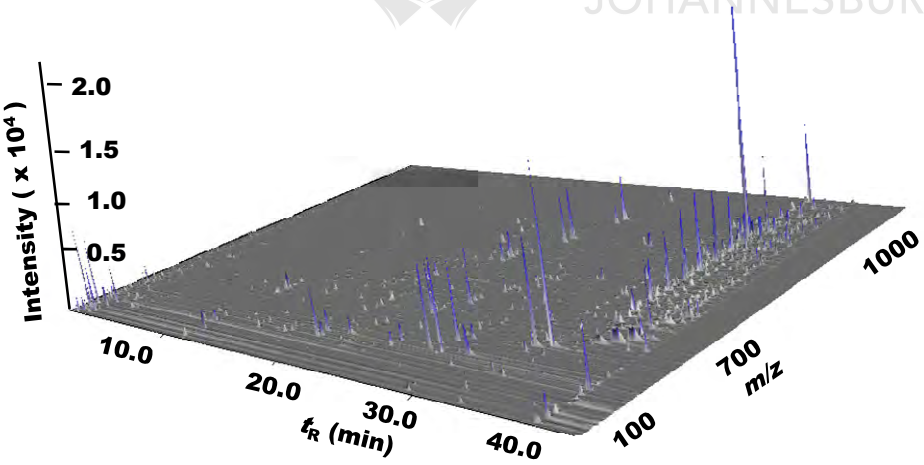
10 min



20 min



40 min



Chapter 4: Collision energy alteration during mass spectrometric acquisition is essential to ensure unbiased metabolomic-based analyses.

4.1. Abstract

Metabolomics entails the identification and quantification of all metabolites within a biological system under a given physiological status and, as such, should be unbiased. Various techniques are employed to measure the metabolite content in living systems, and these differ with the mode of data acquisition and output generation. LC-MS is one of many techniques that has been utilised to study the metabolomes of different organisms but, although used extensively, does not provide a complete metabolic picture. Recent developments in technology, such as the introduction of UHPLC-ESI-MS, have however seen LC-MS as the preferred technique for metabolomics. Here, we show that by varying the MS settings on a UHPLC-ESI-MS, different metabolite profiles result from the same sample. Utilizing the UHPLC-high definition MS instrument, the collision energy was continuously altered (3, 10, 20, and 30 eV), during MS acquisition. PCA and OPLS-DA analyses of the generated UHPLC-MS data of metabolites extracted from elicited tobacco cells revealed different clustering – and distribution patterns. As expected, the ion abundance decreases with a concomitant increase in collision energy levels; but, more importantly, resulted in unique multivariate data patterns from the same samples. Our findings suggest that different collision energy settings should be explored during MS data acquisition as it can contribute to covering a wider percentage of the metabolome by UHPLC-ESI-MS and prevent biased results.

Keywords Metabolomics . Collision energy . 2-Isonitrosoacetophenone . UHPLC-MS . PCA . OPLS-DA.

4.2. Introduction

By virtue of being the end product of cellular metabolism, metabolites act as regulatory components of the proteome, transcriptome and genome, and accumulation thereof allows a complete snapshot of the physiological status of a cell [1-3]. Metabolomics has thus been defined as the comprehensive and quantitative analysis of all metabolites under a given physiological status [4] and the metabolite content of a living organism is referred to as the metabolome [5]. Metabolomics is amongst the fastest growing fields of systems biology and, as such, scientists spend a great deal of time and effort extracting meaningful information from such datasets [6]. Different techniques have been applied to study the metabolite content of different biological systems and, similar to the array of extraction procedures, there is no single analytical technique which is able to cover the complete metabolome in a single analysis due to the heterogeneity thereof [7, 8]. It is thus important to note that different techniques will result in varying metabolite profiles and it is therefore essential to develop a universal method whereby a high percentage coverage of the metabolome can be achieved. As one of the core techniques for metabolomic studies, Ultra-Performance Liquid Chromatography coupled to Mass Spectrometry (UHPLC-MS), allows for both the detection of novel markers and, *via* the provision of elemental composition and fragment ions, biomarker identification.

In the current study, we show that by using UHPLC-MS as a technique of choice, different parameters can affect the metabolomic data output. By altering the collision energy on the trap mode of the Synapt G1 ESI-MS (Waters Corporation, Milford, USA), different metabolite profiles were attained. Here, metabolites extracted from tobacco cell suspensions treated at different time intervals with isonitrosoacetophenone (INAP), a sub-component of a plant-derived stress metabolite with anti-fungal and anti-oxidant properties [9], were analyzed on the UHPLC-MS. The generated data was further compared by unsupervised principal component analysis (PCA), while supervised orthogonal projection to latent structures discriminant analysis (OPLS-DA) models were also applied in order to identify the biomarkers responsible for the separation between various biological groups of samples using different MS settings. From the results it can be seen that the metabolite profiles of the data generated at different MS collision energies varied and thus represents results with unique merit. Data obtained at these different collision energy levels reveal a wider metabolic picture, thus offering more information on the

metabolome under investigation. It is therefore of major importance that, prior to implementing a full scale metabolomics study/analyses, various technical parameters should be considered and validated to ensure an unbiased metabolic coverage.

4.3. Experimental

4.3.1. Chemicals

Isonitrosoacetophenone was purchased from Sigma-Aldrich (St. Louis, MO, USA) and a stock of 250 mM solution in acetone was prepared. Methanol (Romil, Cambridge, UK) was utilised for metabolite extraction and chromatographic separation respectively. Leucine-enkephalin and formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA).

4.3.2. Cell culture, treatment and metabolite extraction

Nicotiana tabacum cv *Samsun* cell suspensions were cultivated as previously described [10] and treated with INAP solubilized in acetone to a final concentration of 1 mM for 6, 12, 18 and 24 h. Following each time interval, metabolites were extracted using 100% methanol, homogenized and centrifuged at 13000 x *g* for 10 min to pellet the cell debris. The supernatants were transferred to new tubes and dried to completeness with the aid of a speed vacuum centrifuge (R.C 10.09, Jouan, France) operating with constant heating at 50ºC. The residual brown precipitates were re-dissolved in 400 µl, 50% (v/v) methanol in water and passed through a 0.22 µm filter using a 1 mL sterile syringe.

4.3.3. Chromatographic analyses

Methanol extracts (5 µl) were analyzed on a UHPLC connected to the Synapt-high definition G1 MS instrument (Waters, Corporation, USA) equipped with an Acquity BEH C18 column (100 mm × 2.1 mm with particle size of 1.7 µm) (Waters Corporation, USA). Two technical replicates for 5 independent samples were performed resulting in 10 injections for each biological group (control, 6, 12, 18, and 24 h). The composition of mobile phase A consisted of 0.1% formic acid in deionized Milli Q water and mobile phase B consisted of 0.1% formic acid in methanol. The column was eluted with a linear gradient at a constant flow rate of 400 µl/min of 5% B over 0.0-

2.0 min, 5-95% B over 2.0-22.0 min, held constant at 95% B over 22.0-25.0 min, 95-5% B over 25.0-27.0 min and a final wash at 5 % B over 27-30 min.

4.3.4. Mass Spectrometry acquisition

The separated analytes were monitored using both photodiode array (PDA) and electrospray ionization mass spectrometry (ESI-MS) detectors. For MS detection, the optimal experimental conditions were as follows: capillary voltage of 2.5 kV, sample cone voltage of 17 V, multichannel plate detector voltage of 1750 V, source temperature of 120°C, desolvation temperature of 400°C, cone gas flow of 50 L/h and desolvation gas flow of 450 L/h. The mass spectrometric full scan data were acquired in both positive and negative ionization mode from 100-1000 Da with a scan time of 0.1 s, inter-scan delay of 0.02 s. Data was centroided and mass spectra corrected in real time by an external reference standard consisting of leucine enkephalin (5 pg/mL) with the lockmass flow rate of 0.4 mL/min and mass window of 0.5 Da. To assist with structure elucidation and compound identification, the MS experiment file was set up to do unfragmented as well as three fragmenting experiments simultaneously. Fragmentation was done at increasing collision energies between 3 to 30 eV on the trap optics of the Synapt MS to obtain substructure information. Unless stated otherwise, unfragmented conditions refer to collision energy level of 3 eV.

4.3.5. Multivariate data analysis

Data collected from the different settings were further analyzed by Markerlynx XS^{TM} software (Waters Corporation, Milford, USA). MarkerLynx XS^{TM} parameters were set to analyze the whole retention (Rt) range of the chromatogram, mass range 100-1000 Da, mass tolerance 0.02 Da and a Rt time window of 0.2 min. Isotopic peaks were excluded from the analysis. Markerlynx XS^{TM} analyses were repeated for all the different data sets representing the various files generated from the altered collision energies. The datasets thus obtained were exported to the SIMCA-P software version 12.0 (Umetrics, Umea, Sweden) programme in order to perform PCA and OPLS-DA. Pareto scaling was used for both models.

4.4. Results and Discussion

As previously stated, there is no single technique which can provide a complete metabolic analysis in a single experiment. NMR for example is regarded as a comprehensive technique, but its lack of sensitivity represents a shortcoming [11]. Different analytical platforms hyphenated to mass spectrometers as detectors provide an alternative to overcome sensitivity shortfalls of NMR [12]. Recently, LC-MS has become the technique of choice for (underivitized) metabolite analyses [13] since it allows molecular identification and quantification of a wide spectrum of differing polarity and neutral metabolites, even when present at relatively low concentration levels in a complex matrix [14]. It allows for both the detection of novel markers and, *via* elemental composition and fragment ions, structural elucidation for biomarker identification.

Varying sensitivity in the acquisition and detection modes from one instrument to another provides an alternative when choosing instruments, and consideration should be made with regard to the ionization mode and analyzer [14]. Depending on the combination of the latter two aspects, it is expected that different metabolite profiles will be obtained. In a previous study, Nordstrom and colleagues [15] has shown that different ionization modes result in distinct metabolite profiles. They also illustrated that data generated with the same ionization mode operating at varying polarities, i.e. ESI (+) or ESI (-), results in significantly different metabolic profiles within the same sample. It is thus from this notion that the current study stemmed, and shows that changes in the collision energy levels on MS acquisition significantly contributes to the detection of a variety of metabolites.

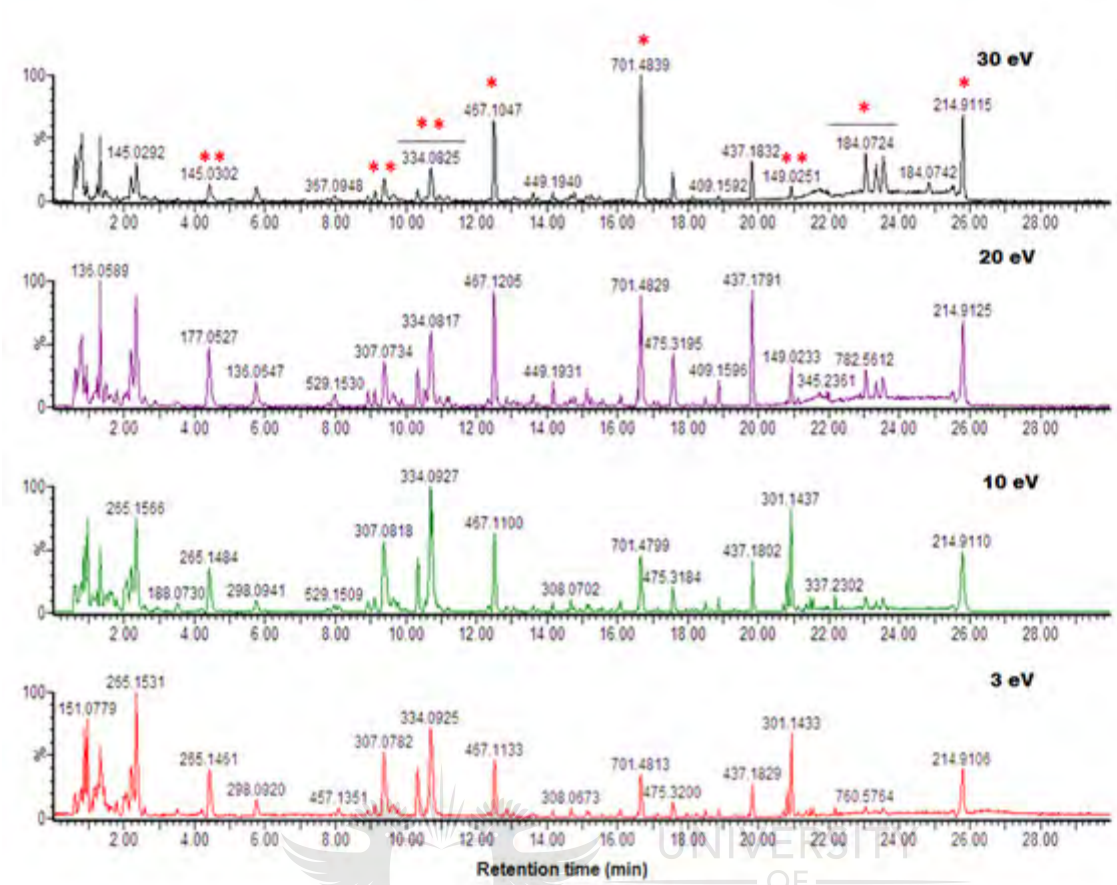


Fig. 1 Overlaid representative, base peak intensity UHPLC-ESI (-)-MS chromatograms, of the 18 h INAP-treated sample extracts analyzed under different collision energy levels (3 – 30 eV). The values above each peak represent the mass thereof. A single asterisk (*) or double asterisks (**) above selected peaks indicate peaks which are showing increasing or decreasing trends respectively with an increase in collision energy. The bar below the stars represents multiple peaks which are showing similar trends of increase or decrease in abundance at different collision energies

PCA has recently found wide application in analyzing metabolomics data [11, 16]. PCA is a mathematical procedure defined as an orthogonal linear transformation of possibly correlated variables into a smaller number of uncorrelated variables called principal components, where greatest variance within the data by any projection is explained on the first co-ordinate (called the first principal component), and the least variance is explained/projected by subsequent principal components [17]. It is such transformation which allows complex data to be visualized

in at least two dimensions, hence simplifying the pattern of the data. PCA was used to evaluate the metabolite patterns of metabolomic data generated from samples analyzed by the same MS operating at different collision energy levels. Taking the 18 h INAP-treated sample extracts for instance, it is visually clear and easy to differentiate UHPLC-MS chromatograms of this sample attained at different collision energies based on the metabolite distribution patterns (Fig. 1). These differences are even more pronounced in the PCA score plots (Fig. 2) where it can be seen that the variation between sample groups is more pronounced along the first component (PC1) at lower energies as compared to the higher ones. In addition, it is also of major importance to note that there is higher variation within the groups along PC 2 associated with an increase in collision energy levels. This observation is an evidence of different metabolite distributions associated with altered collision energy levels.

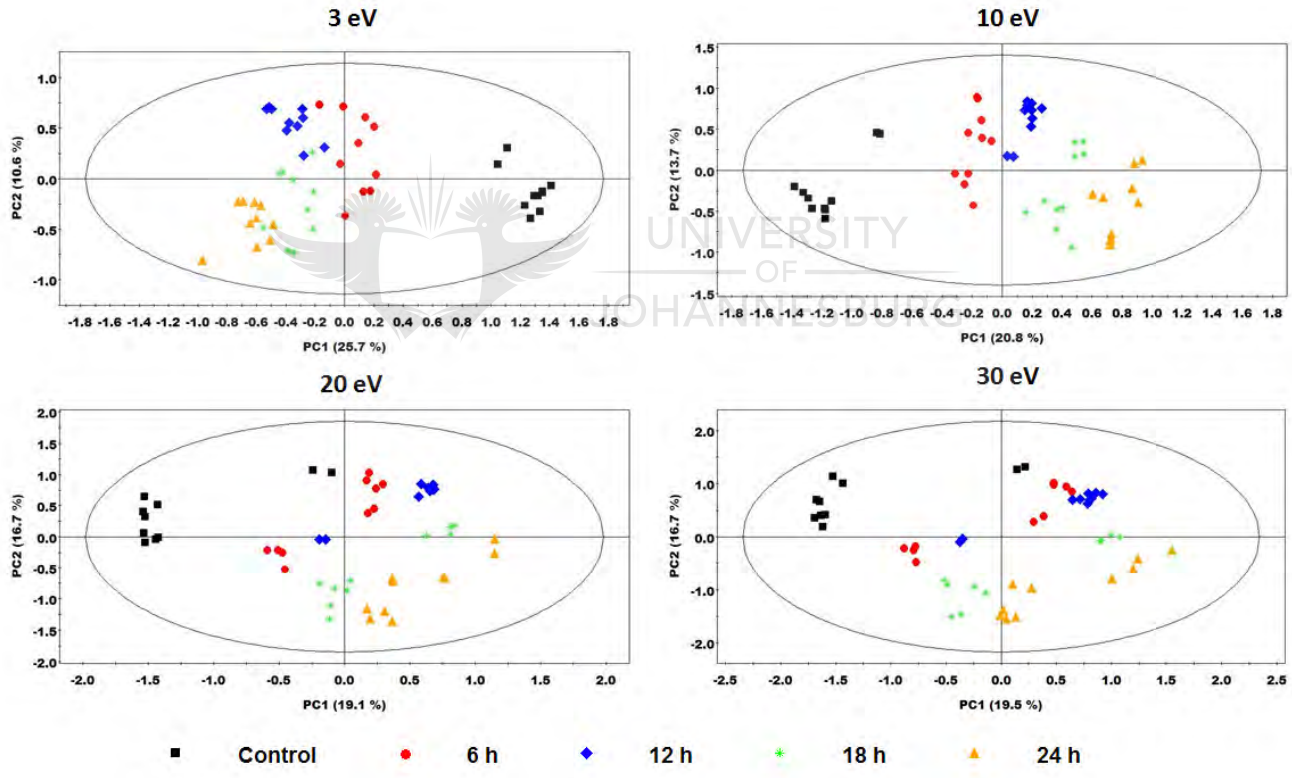


Fig. 2 Representative PCA score plots showing the different clustering/grouping of samples generated using different collision energy levels for each sample

While PCA is an excellent tool for metabolomic data analyses, it does have limitations since separation is obtained only from maximum variations between samples [18]. For the same and

other unidentified reasons, many reports on metabolomic analyses where multivariate statistics are used, also include other models in parallel [18, 19]. Here, orthogonal projection to latent structures discriminant analysis (OPLS-DA), a supervised model, was used as a complimentary model to PCA. The advantage of using OPLS-DA and other supervised models is associated with the ability to highlight biomarkers/variables which are strictly associated with the biological factor under investigation. Using a combination of control and 18 h treated samples, OPLS-DA was employed using two "dummy" classes/identifier - Class 1 for controls and Class 2 for 18 h INAP-treated tobacco cell suspensions. Here the separation between the control and 18 h INAP-treated samples was even more pronounced (Fig. S1, see Electronic supplementary material). The 18 h treatment time point was chosen because it was found to show optimal metabolite accumulation (data not shown). Using the OPLS-DA loading S-plots (Fig. 3) generated from the models indicated in Fig. S1 (see Electronic supplementary material), it can be seen that there is a separation/grouping of controls and INAP-treated samples due to different biomarkers, hence indicating unique metabolic patterns. From the same figure, it can also be noted that more potential biomarkers were generated when lower collision energy was used. In order to simplify/highlight the above observation, the top 20 biomarkers/variables projected along the side of the predictive component (p1) of the loadings S-plot, representing the treated samples, were compared. A table containing these biomarkers was constructed and the unique metabolite ions generated from a particular collision energy are highlighted (Table S1, see Electronic supplementary material). From the above observations, it is thus convincing that metabolomic data output from an MS operating at alternate collision energy levels differ significantly.

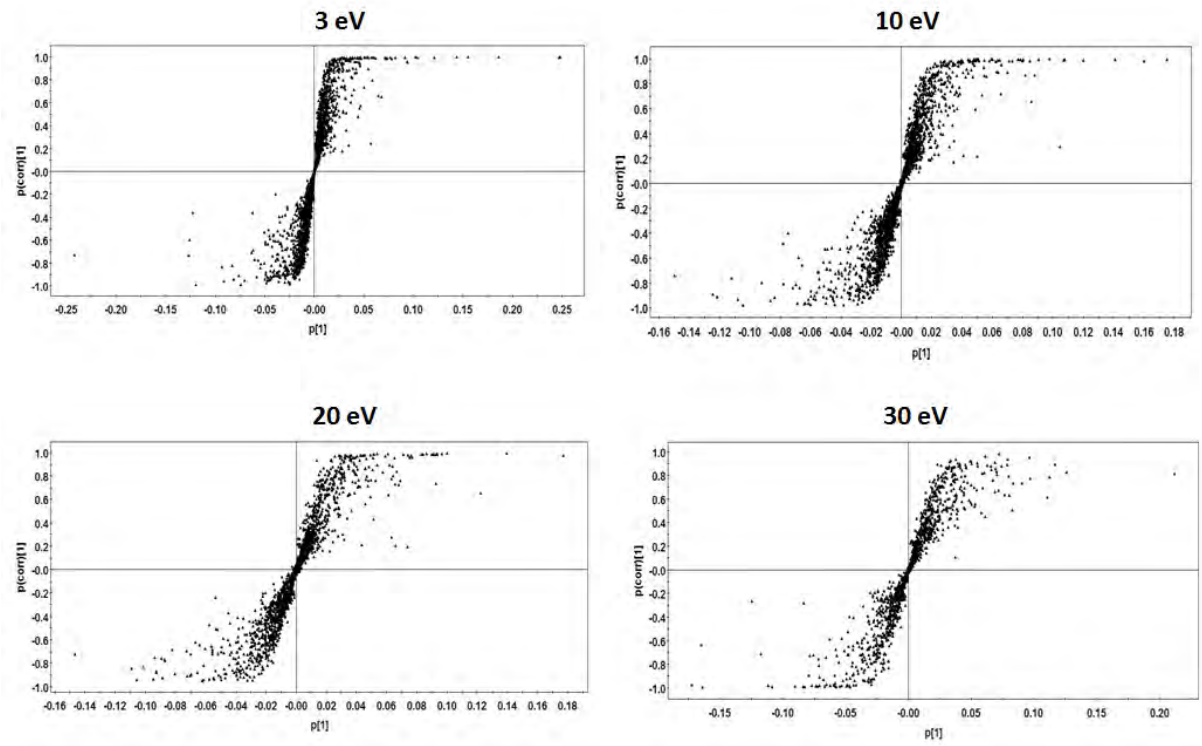


Fig. 3 Representative OPLS-DA loadings S-plots showing the distribution patterns of ions contributing to the differences between control and 18 h INAP-treated samples at different collision energy levels on the score plots (Fig. S1)

In addition to the two figures provided as evidence to show the effect of different collision energy, additional comprehensive data analyses were also carried out, including the fragmentation behavior of single ions at different collision energy levels [20]. It was observed that, by changing the fragmentation energy, more structural features can be deduced from the MS data as compared to when a single energy level is used (e.g. Electronic supplementary material, Figs. S2 A and B for scopoletin and naringenin).

4.5. Conclusion

Changes in the metabolite concentrations define the consequences of cellular processes, and hence allow ultimate evaluation of the physiological status of the cell/organism. However, in order to gain this information from a global view of the metabolome, analytical strategies should be well optimized to remove artifacts and meet high throughput demands with the aim of gaining comprehensive metabolite profiles. It is therefore important that analytical techniques are well

validated and optimized for the latter to be possible. The data presented in this work reveal how metabolite profiles/patterns can be influenced and optimized through careful consideration of mass spectrometry acquisition parameters such as collision energy. By changing the collision energy levels, it was found that different patterns of metabolite distribution are achieved. When the unfragmented (3 eV) conditions were utilized, the metabolite patterns and yields differed significantly from those which were attained using 10 eV, 20 eV or 30 eV. Varying collision energies are used to assist in structural elucidation, but should be properly monitored and reported so as to maintain consistency with data acquisition in order to attain quality and consistent metabolomic data. Higher collision energies will generate more complex data sets that may be able to distinguish between different control and treatment groups, but could also lead to more ambiguous results if wrongly interpreted. Care must therefore be taken when selecting an appropriate energy level that will allow maximum data output and sufficient structural elucidation of affected metabolites prior to holistic metabolomic studies. The results also support the call by the metabolomics community to honor the guidelines for reporting metabolomics data to ensure consistency in data generating platforms.

4.6. References

1. Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Nat Rev Mol Cell Biol 5: 763–769

2. Ryan D, Robards K (2006) Anal Chem 78: 7954–7958

3. Seger C, Sturm S (2007) J Prot Res 6: 480–497

4. Fiehn O (2001) Comp Func Genom 2: 155–168

5. Oliver SG, Winson MK, Kell DB, Baganz F (1998) Trends Biotechnol 16: 373–86.

6. Robertson DG (2005) Toxicol Sci 85: 809-822

7. Oksman-Caldentey K-M, Saito K (2005) Curr Opin Biotech 16: 174-179

8. Lu X, Zhao X, Bai C, Zhao C, Lu G, Xu G (2008) J Chrom B 866: 64-76

9. Dubery IA, Louw AE, Van Heerden FR (1999) Phytochemistry 50: 983-989

10. Sanabria NM, Dubery IA (2006) Biochem Biophys Res Comm 344: 1001-1007

11. Kim HK, Choi YH, Verpoorte R (2010) Nat Protoc 5: 536-549

12. Dunn WB, Bailey NJC, Johnson HE (2005) Analyst 130: 606-625

13. Grata E, Boccard J, Guillarme D, Glauser G, Carrupt PA, Farmer EE, Wolfender J-L, Rudaz S (2008) J Chrom B 871: 261-270

14. Villas-Boas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J (2005) Mass Spectrom Rev 24: 613-646

15. Nordstrom A, Want E, Northen T, Lehtio J, Siuzdak G (2008) Reprod Biol 80: 421-429

16. Liu NQ, Cao M, Frédérich M, Choi Y-H, Verpoorte R, Van der Kooy F (2010) J Ethnopharmacol 128: 230-235

17. Jolliffe IT (2002) Principal Component Analysis, 2nd ed. Springer, New York

18. Dao TTH, Puig RC, Kim HK, Erkelens C, Lefeber AWM, Linthorst HJM, Choi Y-H, Verpoorte R (2009) Plant Physiol Biochem 47: 146-152

19. Maree JE, Viljoen AM (2011) Vibrational Spec 55: 146-152

20. Plumb RS, Johnson KA, Rainville P, Smith BW, Wilson ID, Castro-Perez JM, Nicholson JK (2006) Rapid Commun Mass Spec 20: 1989-1994.

4.7. Supplementary files

Electronic supplementary Table S1

List of the top 20 most discriminative biomarkers derived from the OPLS-DA S-loadings plot of the data analyzed at different collision energies. The biomarkers are identified by the respective Rt and *m/z* values as in (retention time)_(base peak mass).

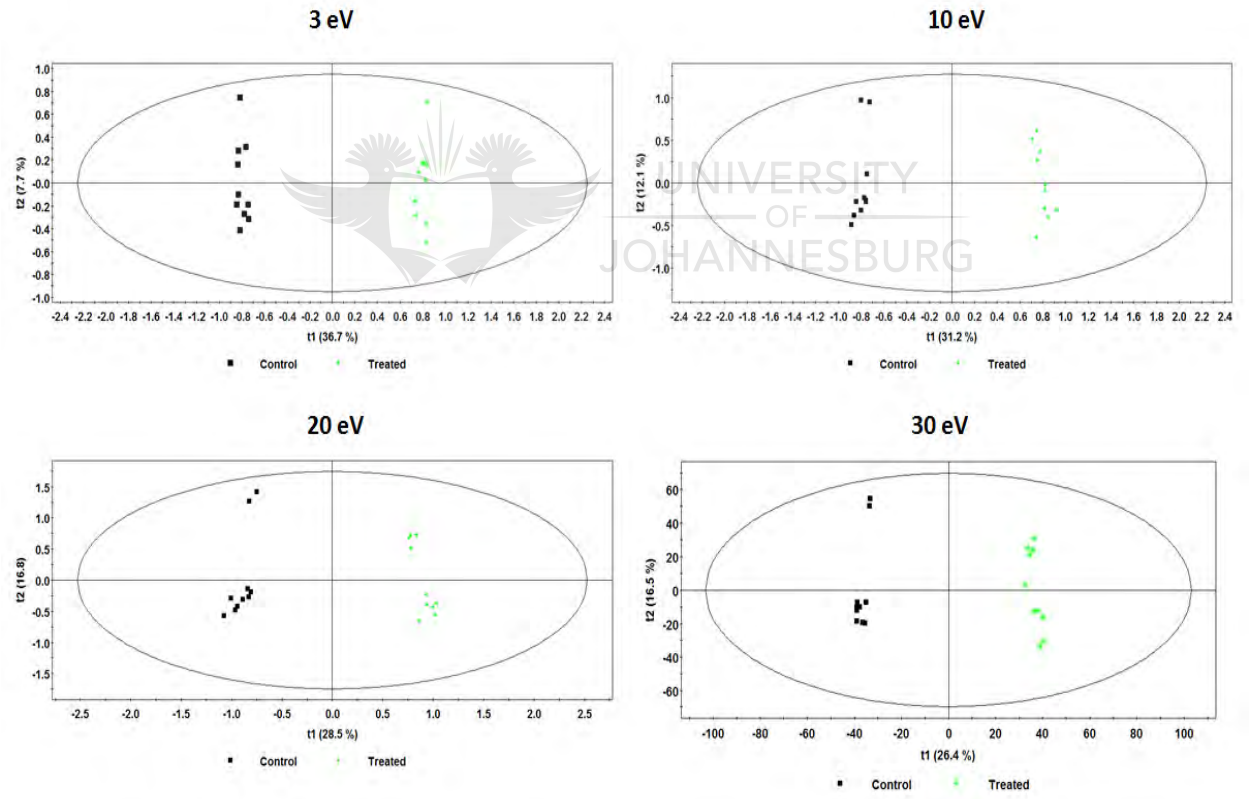
Unfragmented	**10eV**	**20eV**	**30eV**
1.36_194.9460	1.24_331.9483	10.32_180.0137	1.36_194.9474
1.37_176.9353	1.37_176.9360	10.32_306.0775	1.36_292.9231
10.33_453.1064	10.33_272.0894	10.33_272.0899	10.32_328.0597
10.70_179.0557	10.33_306.0776	10.33_328.0596	10.33_180.0138
10.70_346.0699	10.33_453.1073	10.70_674.1027	10.70_172.9545
10.70_356.0976	10.67_179.0554	11.18_371.0995	12.50_249.0621
10.70_357.1039	10.70_356.0992	12.49_371.0973	12.50_371.0983
10.95_149.0255	10.75_455.0144	12.50_249.0619	12.50_372.1016
11.18_371.0983	11.18_371.0980	12.50_372.1016	12.50_393.0802
12.49_371.0980	12.49_371.0973	16.66_677.4961	16.66_677.4959
8.09_455.1239	12.50_372.1015	17.41_188.0187	16.67_678.5003
8.92_491.1395	16.67_723.5010	23.12_265.1480	17.41_188.0188
9.38_207.0510	17.41_188.0183	25.78_304.9124	23.12_265.1486
9.38_284.0873	25.79_304.9124	25.78_434.8702	25.78_304.9123
9.38_319.0592	8.92_445.1341	9.38_376.0104	25.78_434.8705
9.38_330.0927	9.38_283.0827	9.38_498.0472	25.78_564.8275
9.39_121.0304	9.38_329.0879	9.38_620.0821	9.38_121.0310
9.39_283.0810	9.38_620.0826	9.39_121.0309	9.38_376.0095
9.39_329.0865	9.39_121.0304	9.63_272.0893	9.38_498.0478
9.63_599.1650	9.63_599.1643	9.63_306.0774	9.64_328.0593

The table highlights discriminatory metabolites (given by Rt and *m/z*), thus metabolites of which the abundance/distribution are responsible for the separation between control and treated (18 h)

sample datasets acquired at the different collision energy levels. The ions are arranged in such a way that the most influential ions are on top, whilst the bottom ones represents those with less influence. From the table it is quite notable that different ions are responsible for the separation of samples at different collision energy levels.

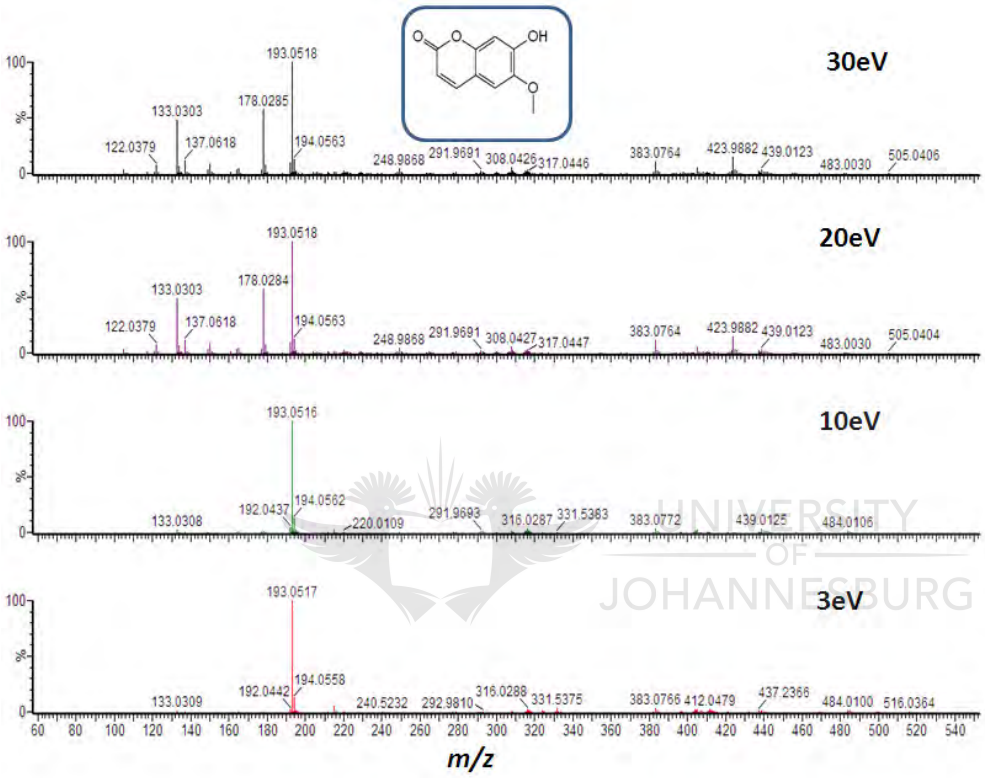
Electronic supplementary Fig. S1

Representative OPLS-DA score plots showing the different clustering/grouping of INAP-treated (18 h) cell extract samples generated using different collision energy levels. Plots were generated where t1 and t2 represent the predictive component and orthogonal component, respectively.



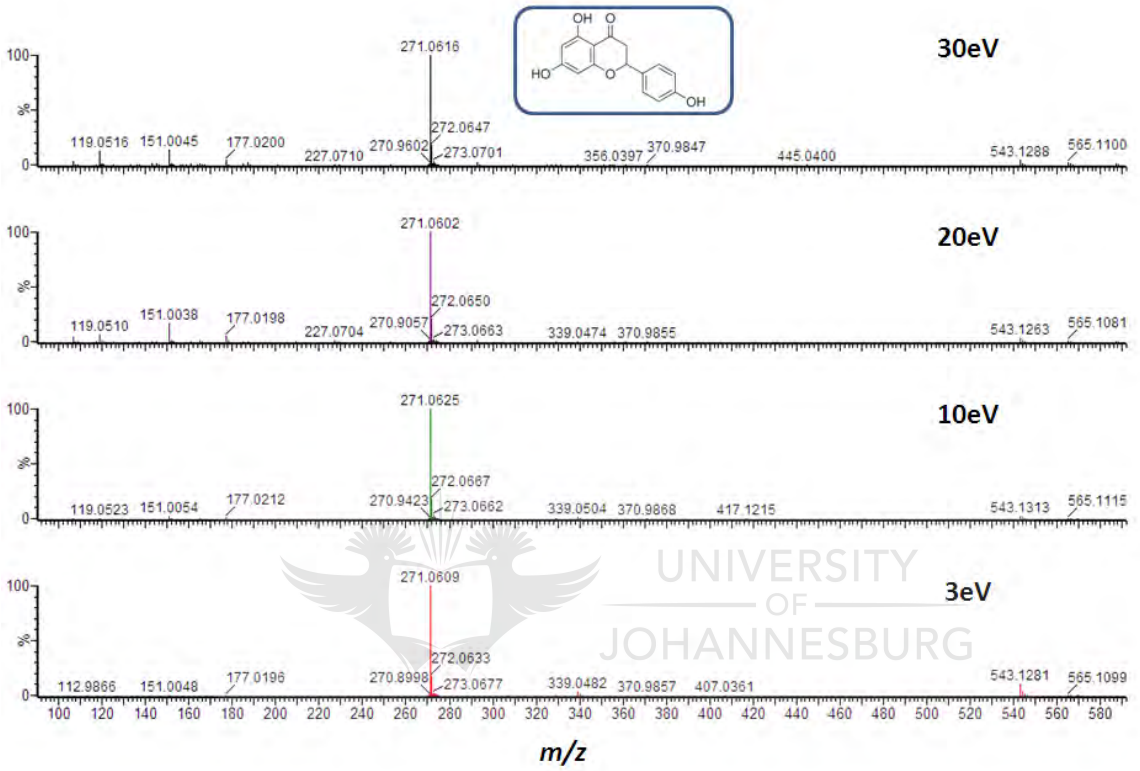
Electronic supplementary Fig. S2 A

Overlaid MS spectra showing differential fragmentation patterns of scopoletin at different collision energy levels



Electronic supplementary Fig. S2 B

Overlaid MS spectra showing differential fragmentation patterns of naringenin at different collision energy levels



From the results (Electronic supplementary Figs. S2 A and B), it is clear that changes in collision energy levels affect the fragmentation data detected by mass spectrometry. However, all metabolomic studies have the common intention which is to decipher the biological meaning of the attained response. This is mainly achieved by revealing the identity of metabolites of which the accumulation is perturbed by the treatment in question. By so doing, the affected pathways can further be elucidated and studied in more depth and detail. Here, scopoletin (a coumarin), was used as authentic standard to evaluate its structural stability under different collision energy levels. There is an increase in mass ions lower than the pseudomolecular ion in the MS spectrum with increasing collision energy levels (Fig. S2 A), hence more complex spectral data relating to structural features of the molecule. However, higher collision energy (30 eV) proved to be

detrimental to the process and resulted in extensive loss of spectral quality (and quantity) and masked the structural identity. The above was proven by matching the attained raw data with those of a custom library at the CSIR (Pretoria, South Africa). It was seen that at a collision energy of 3 eV, the percentage probability hit was 58%, for 10 eV 89.8%, for 20 eV 96%, and dropped to only 21% at 30 eV.

The same differential fragmentation pattern was also observed in several peaks from the raw data, especially those which showed a high degree of variability in abundance/intensity across different collision energy levels (Fig. 1) and in most cases 10-20 eV collision energy levels were found to be moderate for structural elucidation. It should, however, be stressed that not all metabolites will show this trend. Fig. S2 B shows the fragmentation pattern of the flavonoid, naringenin. Here it can be seen that this molecule showed a more conserved fragmentation pattern across all levels of collision energy when compared to scopoletin. This observation suggests that changes in collision energy does not only result in more complex data output but also affect the detailed structural information of metabolites under investigation.

The collection of data from multiple simultaneous MS experiments can therefore lead to large data sets that can be mined to obtain mass spectra of unfragmented compounds by keeping the collision energy of the first experiment below 10 eV. Another experiment at 10 – 20 eV would reveal low to moderate fragmentation and supply additional information to aid with structure elucidation. Increasing the collision energy to 30 – 40 eV would result in severe fragmentation of the more labile molecules, while the more robust molecules will only start to show low to moderate fragmentation.

Chapter 5: Isonitrosoacetophenone induces perturbations in the metabolic status of tobacco cells.

5.1. Abstract

Plants have developed biochemical and molecular responses to adapt to different stress environments. One of the characteristics of the multi-component defence response is the production of defence-related metabolites. Plant defences can be triggered by various stimuli, including synthetic or naturally occurring molecules, especially those derived from pathogens. In the current study, *Nicotiana tabacum* cell suspensions were treated with isonitrosoacetophenone (INAP), a sub-component of a plant-derived stress metabolite with anti-fungal and anti-oxidant properties, in order to investigate the effect thereof on cellular metabolism. Subsequent metabolomic-based analyses were employed to evaluate changes in the metabolome. UHPLC-MS in conjunction with multivariate data analyses was found to be an appropriate approach to study the effect of chemical inducers like INAP on plant metabolism in this model system. Principal component analysis (PCA) indicated that INAP is capable of inducing time-dependent metabolic perturbations in the cultured cells. Orthogonal projection to latent structures disciminant analysis (OPLS-DA) revealed metabolites of which the levels are affected by INAP, and eight of these were identified from the mass spectral data and online databases. These metabolites are known in the context of plant stress– and defense responses and include benzoic- or cinnamic acid derivatives that are either glycosylated or quinilated as well as flavonoid derivatives. The results indicate that INAP affects the shikimate - , phenylpropanoid - and flavonoid pathways, the products of which may subsequently lead to an anti-oxidant environment *in vivo*.

***Keywords*:** Biotransformation; induced-defences; isonitrosoacetophenone; 2-keto-2-phenyl-acetaldoxime; metabolism; metabolomics; multivariate statistics; *Nicotiana tabacum;* xenobiotics.

5.2. Introduction

Plants can specifically recognize pathogenic micro-organisms and respond by activating appropriate multi-component defence mechanisms (McDowell and Dangl, 2000; Odjakova and Hadjiinova, 2001). By virtue of not having an adaptive immunity, plants rely solely on innate immunity of which there are two types namely induced - and preformed immunity. Inducible plant immunity is the most complex of these, involving the generation of reactive oxygen species (ROS), defence gene activation, activation of programmed hypersensitive cell death responses (HR), and the production of pathogenesis-related (PR) proteins and anti-microbial phytoalexins (Dixon *et al*., 1994; Van Loon, 1997). Cell wall strengthening adds to existing structural defenses, thereby restricting entry of the pathogen to avoid further infection (Hammond-Kosack and Jones, 1996; Greenberg, 1997; De Ascensao and Dubery, 2000; Zwiegelaar and Dubery, 2006). Chemical defence is regarded as a second line of defence after structural modifications and is characterised by the production of chemical compounds with direct/indirect anti-microbial activities. Most of these anti-microbial metabolites include, amongst many, phenolic compounds, terpenoids, cyanogenic glycosides, hydroxamic acids and peptides (Bell, 1981; Bednarek and Osbourn, 2009). Since metabolites accumulate as the end products of cellular metabolism, and the levels thus reflect the organism's ultimate response to biological or environmental changes (Fiehn, 2002).

In addition to / as a consequence of innate immune responses, resistance can be induced in plants. The significance of the HR response is to avoid further spreading of the attacking pathogen to other uninfected sites. Moreover, following the HR, uninfected sites become more resistant to subsequent infection, an adaptive phenomenon known as systemic acquired resistance (SAR) (Ryals *et al*., 1996). More recently, chemicals have also been employed to trigger a condition analogous to SAR, the most widely used being benzothiadiazole (BTH) with the trade name BION (Gatz and Lenk, 1998; Oostendorp *et al*., 2001; Dao *et al*., 2009). Promoters of genes with direct activity towards plant defence responses have been shown to respond to different types of chemical inducers (Gorlach *et al*., 1996; Gatz, 1997). Although BTH is the most researched chemical inducer of plant defence, other molecules are also known to exhibit this activity. These include, amongst others, β-aminobutyric acid (BABA) (Jakab *et al*., 2001), methyl-2,6-dichloroisonicotinic acid (INA) (Metraux *et al*., 1991), azelaic acid (Jung

et al., 2009) and more recently, riboflavin (Liu *et al*., 2010b). The mechanism of action of these molecules is in most cases not well documented.

Dubery *et al.* (1999) reported the accumulation of an oxime-containing stress metabolite/phytoalexin, (4-(3-methyl-2-butenoxy)-isonitrosoacetophenone or citaldoxime) in citrus peel undergoing oxidative stress due to gamma radiation treatment. This novel compound was reported to exhibit phytoalexin, anti-oxidant and radical scavenging activities (Dubery *et al*., 1999). Although oxime functional groups are rare in natural products, they occur in a variety of phyla, e.g., sponges, bacteria, fungi, and plants (Almeida *et al*., 2011). In plants, oximes are known to be intermediates of a range of metabolic pathways (e.g. nitriles, cyanogenic glycosides, glucosinolates *etc.*) subject to controls that result in variation in both the type and amount of end product formed (Mahandevan, 1973). In the context of plant defence/stress responses, aldoximes are intermediates/precursors during the biosynthesis of glucosinolates and cyanogenic glycosides, two classes of molecules that play vital roles during plant:herbivore interactions (Moller, 2010).

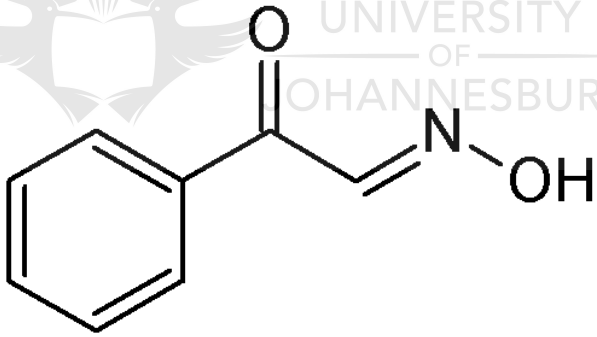


Fig. 1. The chemical structure of isonitrosoacetophenone (INAP).

In the current study, isonitrosoacetophenone (INAP, or 2-keto-2-phenyl-acetaldoxime) (Fig. 1), a compound structurally similar to 4-(3-methyl-2-butenoxy)-isonitrosoacetophenone, was used to investigate induced metabolic changes in tobacco cell suspensions. Ultra-performance liquid chromatography coupled to mass spectrometry (UHPLC-MS) was used for measuring the levels of physiological metabolites affected by INAP treatment, and principal component analysis (PCA) and orthogonal projection to latent structures disciminant analysis (OPLS-DA) of the UHPLC-MS data discriminated between the metabolite content of untreated (control) and

INAP-treated tobacco cell suspensions. The findings presented here are the first in which the effect of INAP on the cellular metabolism of a plant model is investigated.

5.3. Experimental

5.3.1. Cell treatment, extraction and sample preparation

Nicotiana tabacum cv. *Samsun* cell suspensions were cultivated as previously described (Sanabria and Dubery, 2006). Three days after subculturing, aliquots (0.2 g cells/mL suspension) were treated with a 250 mM stock solution of INAP dissolved in acetone, to a final concentration of 1 mM with continuous rotation at 80 rpm and 25ºC. The final concentration of acetone was 0.4 %. Control cells received no treatment.

The experimental design included three biological replicates with five technical repeats in order to have maximum reproducibility of the data. After elicitation time intervals of 6, 12, 18 and 24 h, cells were collected by means of centrifugation at 10 000 x *g* for 10 min in a microcentrifuge at room temperature. Pelleted cells were homogenised in 1 mL 100% methanol using an Ultraturrax rotating blade homogeniser to terminate any enzymatic activity. Extraction of metabolites was maximised by means of further heating the homogenate at 60ºC for 10 min followed by mild sonification for 20 min. The homogenates were centrifuged at 10 000 x *g* for 10 min, the resulting supernatants transferred to new tubes and dried to completeness using a speed vacuum centrifuge operating with constant heating at 50ºC. The residual brown precipitates were re-dissolved in 400 µL, 50% (v/v) methanol in water and filtered through a 0.22 µm filter using a 1 mL sterile syringe. The filtrates were transferred to glass vials fitted with 500 µL inserts and capped. Unless stated elsewhere, all reagents and solvents were of UHPLC grade.

5.3.2. Total phenolic acid and antioxidant determination

Total soluble phenolics were measured using the Folin-Ciocalteau assay (De Ascenao and Dubery, 2003). Here, the extracts (50 µL) were diluted in 50% methanol (0.25 mL) and half-strength Folin-Ciocalteau reagent (0.25 mL), and mixed thoroughly. After mixing, 0.5 mL of saturated aqueous sodium carbonate was added, the mixture again mixed thoroughly and further incubated in room temperature for 1 h. Absorbances were measured at 600 nm against 50%

methanol as a blank. In order to evaluate/screen for the antioxidant molecules induced by INAP treatment, extracts (20 µL) were separated on silica gel 60 F_{254}TLC plates. The plates were developed using chloroform: methanol (60:40 v/v) as mobile phase and dried at room temperature. The TLC plates were sprayed with 1 mM of diphenylpicryl-hydrazyl (DPPH) in methanol and incubated at room temperature for at least 1 h (Esterhuizen *et al.*, 2006). After incubation, the areas showing molecules with antioxidant activity appeared as white zones against a dark purple background of DPPH.

5.3.3. Ultra-performance liquid chromatography-mass spectrometry

The methanol extracts (5 µL) were analyzed on a Waters UHPLC connected to high definition ion mobility MS instrument (UHPLC-QTOF SYNAPT G1 HD-MS system) equipped with an Acquity BEH C18 column (100 mm × 2.1 mm with particle size of 1.7 µm) (Waters Corporation, Milford, MA, USA). The composition of mobile phase A consisted of 0.1% formic acid in deionized water and mobile phase B consisted of 0.1% formic acid in methanol. The column was eluted with a linear gradient at a constant flow rate of 400 µL/min of 5% B over 0.0-2.0 min, 5-95% B over 2.0-22.0 min, held constant at 95% B over 22.0-25.0 min, 95-5% B over 25.0-27.0 min and final wash at 5% B over 27-30 min.

The separated analytes were monitored using both photo-diode array (PDA) and electro-spray ionisation mass spectrometry (ESI-MS) detectors. For PDA detection, the UHPLC LG (Waters Corporation, Milford, MA, USA) detector was used. The sampling rate was set at 20 points/sec; the filter time constant was 0.1 sec. The range of detection was set between 220-500 nm with the resolution of 1.2 nm.

For MS detection, experimental conditions were as follows: ESI (+ & -) conditions: ES capillary voltage: 2.5 kV, sample cone voltage: 17 V, MCP detector voltage: 1600 V, source temperature: 120°C, desolvation temperature: 350°C, cone gas flow: 50 L/h, desolvation gas flow: 450 L/h, m/z range: 100-1000, scan time: 0.1 sec, interscan delay: 0.02 sec, mode: centroid, lockmass: leucine enkephalin (556.3 g/mol), lockmass flow rate: 0.4 ml/min, mass accuracy window: 0.5 Da. The molecular formula assignments software used was MassLynx XS^{TM} (Waters Corporation, Milford, MA, USA). Mass accuracy of all the *m/z* values in all the acquired spectra were automatically corrected during acquisition based on calibration curves,

lockmass and dynamic range extended. The MS detector was also operated in the positive ion mode but there were no significant changes to the results.

5.3.4. Data analyses

MassLynx version 4.1 software (Waters Corporation, Milford, MA, USA) with an added statistical programme for multivariate data analysis, was used to analyse the UHPLC-ESI-MS data where ESI negative raw data was extracted and analysed. MarkerLynx^{TM} parameters were set to analyse the whole retention (t_R) range of the chromatograms, mass range 100-1000 Da, mass tolerance 0.02 Da and a t_R time window of 0.2 min. For qualitative visualisation, isotopic peaks were excluded from the analysis but included for quantitative and identification purposes. The dataset obtained from MarkerLynx^{TM} processing was exported to the SIMCA-P software version 12.0 (Umetrics, Umea, Sweden) programme in order to perform PCA and OPLS-DA models, and Pareto scaling was used for both models. PCA is a non-supervised mathematical procedure which reduces the dimensionality of data without altering the data itself. By mathematical definition, this process is defined as an orthogonal linear transformation of possibly correlated variables into a smaller number of uncorrelated variables called principal components, where greatest variance within the data by any projection is explained on the first co-ordinate (called the first principal component), and the least variance is explained/projected by subsequent principal components (Jolliffe, 2002; Liu *et al*., 2010a). OPLS-DA, on the other hand, is a linear regression method that has been employed successfully for prediction modeling in various biological and biochemical applications (Trygg and Wold, 2002; Cloarec *et al*., 2005; Maree and Viljoen, 2011) and filters out any variation that is not directly related to the response, thereby resulting in models which are easier to interpret (Liu *et al*., 2010a). Distribution graphs and statistical tests were done with the aid of Minitab, version 15 statistical software and n-Anova using Microsoft Excel software, respectively.

Both PCA and OPLS-DA score plots were used to depict the clustering of biological groups depending on the background/source, i.e. either the data originated from control or treated samples. Other plots such as loadings- and S-plots were also utilised for identification of biomarkers of which the levels were affected due to INAP treatment. It is important to note that the peak area on the chromatograms corresponding to residual, non-metabolised INAP, was

removed prior to all statistical analyses. For metabolite identification, Markerlynx results were exported to the Taverna workbench (**http://www.taverna.org.uk**) for PUTMEDID_LC-MS workflow (Brown *et al.,* 2011). These workflows have three main functions which are: correlation analysis, metabolic feature annotation, and metabolite annotation. All these functions allow for peaks which share similar features, such as t_R, to be grouped together and further allows grouping together and annotating features with the type of *m/z* ion (isotope, adduct, dimer, others) which are believed to originate from the same metabolite. Once the aforementioned correlation is made, the elemental composition/molecular formula of each *m/z* ion is automatically calculated and lastly compared to pre-defined databases or manually searched against freely online databases such as DNP, Chemspider, AraCyC, PlantCYC, KEGG, LipidMAPS, KNApSAcK and METLIN database.

5.4. Results and Discussion.

Oxime functional groups are rare in natural products; in plants, oximes are intermediates of a range of metabolic pathways subject to controls that result in variation in both the type and amount of end product formed (Mahadevan, 1973). As noted, the biological roles of oxime compounds such as citaldoxime in plants are not well researched with regard to biosynthesis, interconversions and biological effects (Madala *et al*., 2012). Here, we report on the metabolic changes in cultured *N. tabacum* cells due to INAP treatment, and the current study lays a good foundation for investigating the roles of other oxime-containing molecules in non-cyanogenic plants such as *Nicotiana* or *Solananum* species.

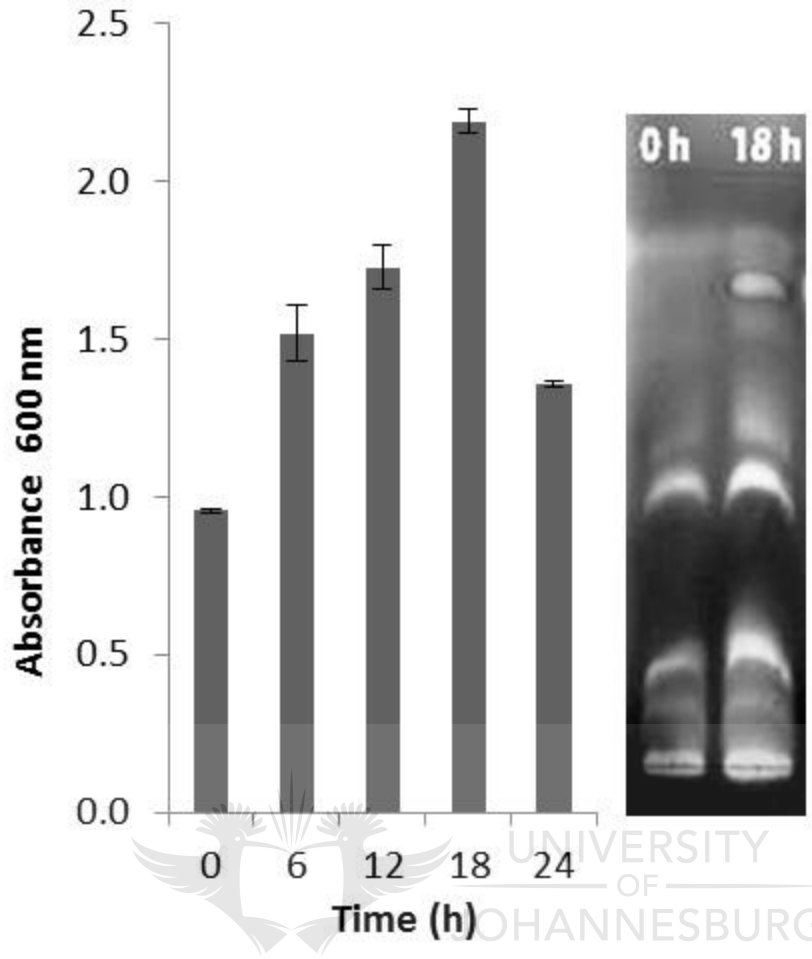


Fig. 2. Graphical representation of time-dependent (0-24 h) accumulation of total phenolics in INAP-treated tobacco cells. The insert is a TLC-DPPH assay showing the separation of antioxidant molecules appearing as white spots against a gray background.

N. tabacum cells responded to treatment with 1 mM INAP by enhanced synthesis of phenolic compounds as detected by the Folin-Ciocalteau reagent (Fig. 2). High resolution chromatography of cellular methanol extracts revealed that INAP treatment affected the metabolome and the changes in the metabolite profiles over different treatment times are easily noted by comparison of the UHPLC-PDA chromatograms (Fig. 3). In addition to the peak corresponding to residual INAP, two dominant peaks were found to exist predominantly only in INAP-treated samples as compared to control samples. PCA and OPLS-DA models were also used to identify the areas of the chromatograms which are responsible for differences between

the control and treated samples, hence identification of endogenous and INAP-derived metabolites of which the levels are affected due to the treatment.

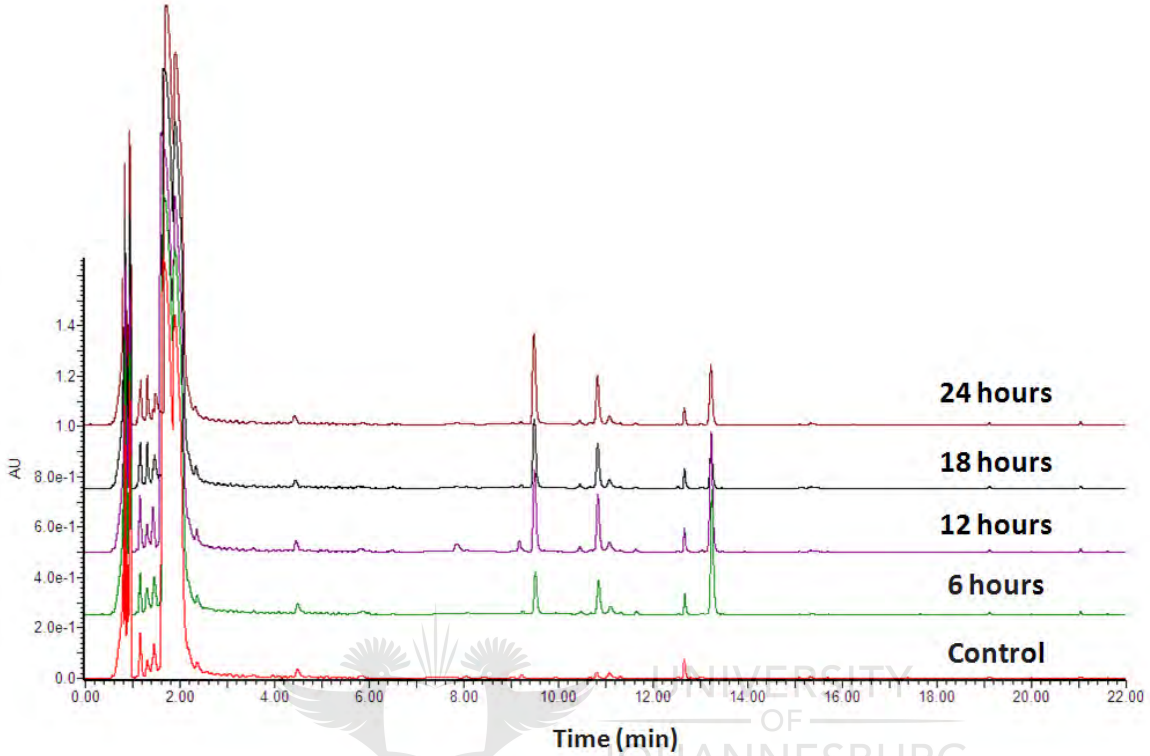


Fig. 3. Comparison of representative UHPLC-PDA base peak intensity (BPI) chromatograms of tobacco cell suspension samples treated with INAP for different time intervals.

Fig. 4A shows the PCA score plot based on UHPLC-MS chromatograms of the control and INAP-treated tobacco cell suspensions (6, 12, 18 and 24 h post treatment). It can be seen that clustering based on different treatment times was achieved. The separation between the controls and treated cells at different treatment time intervals is projected along the first principal component (PC1) which explains the high percentage (32.7%) of the variation in the model. The variation within the group is projected along the second principal component (PC2) which amounts to the least percentage of variation in the model (8.09%). Looking at the separation within the groups of the same treatment along PC2, it can be observed that there is very little variation. Other than the two peaks which are dominant across all the INAP-treated samples, the PCA loadings plot reveals other areas of the chromatograms representing metabolites

(biomarkers) of which levels are affected by this treatment (outliers in the four quadrants of Fig. 4B).

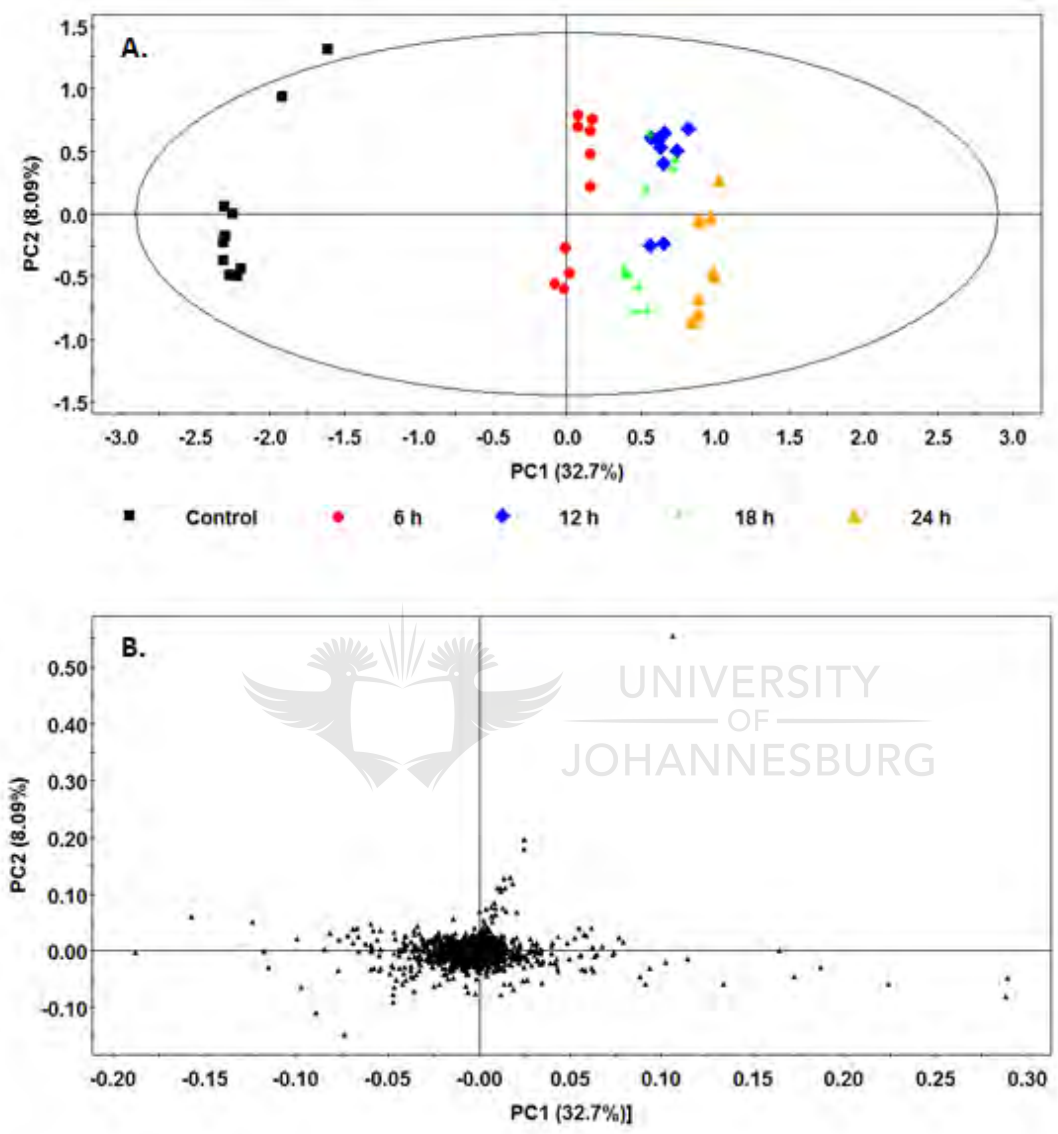


Fig. 4. Representative PCA score plot (A), based on the UHPLC-MS chromatograms, showing clustering of samples from tobacco cell suspensions treated with INAP for different time intervals and its corresponding loadings plot (B), showing biomarkers responsible for the clustering observed in A. The ellipse represents Hotelling's T2 with 95% confidence. Model validation gave R^2X (0.486) and Q^2 (cum) (0.351) with a 4 PC model.

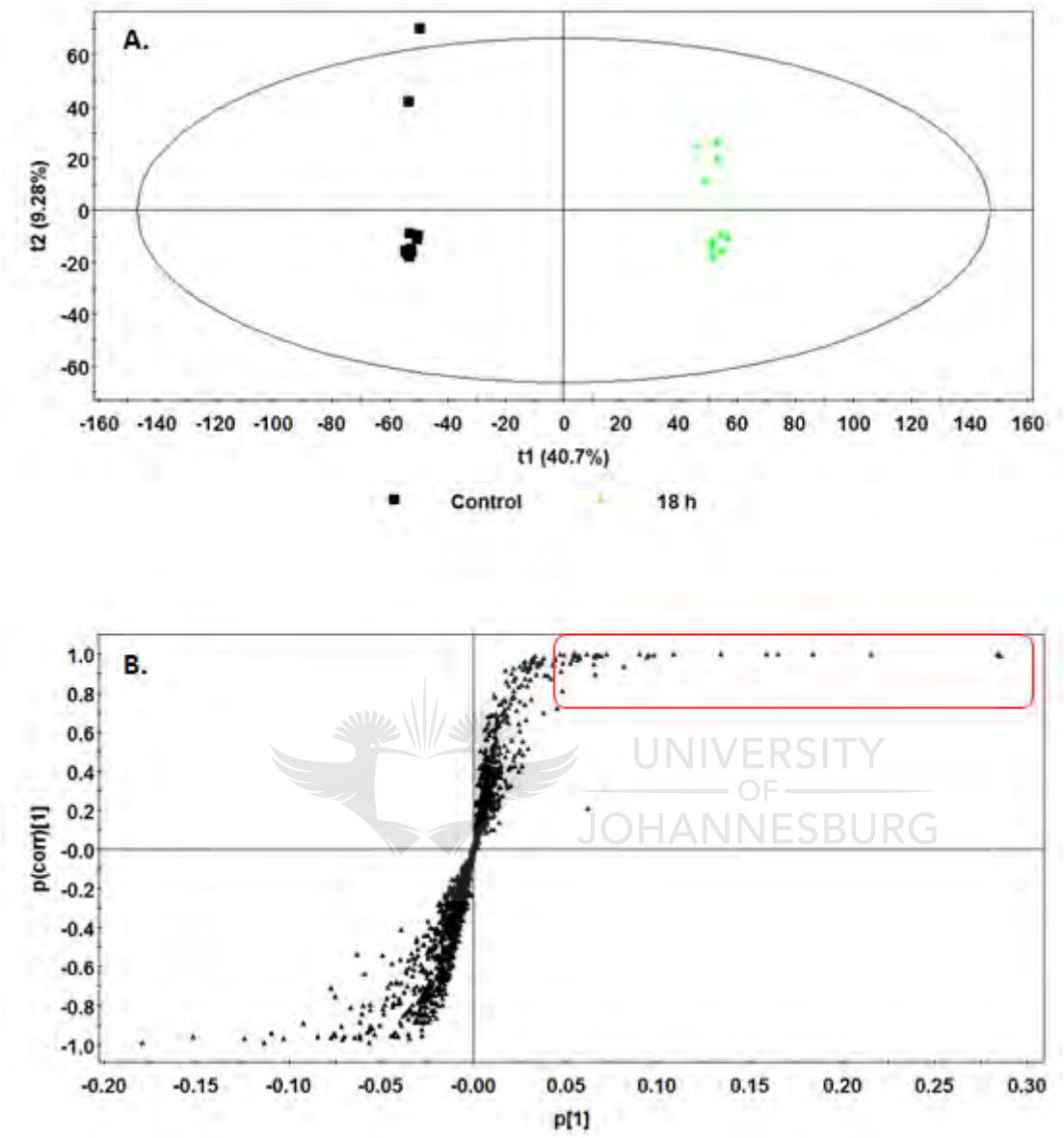


Fig. 5. Representative OPLS-DA score plot (A), based on the UHPLC-MS chromatograms, showing clustering of control *vs*. 18 h treatment of tobacco cell suspensions with INAP and its corresponding loadings S-plot (B), and biomarkers responsible for the clustering observed in A, with those most significantly contributing to the treatment response highlighted by the rectangle. The ellipse represents Hotelling's T2 with 95% confidence. Model validation gave R^2X (0.500), R^2Y (0.998) and $Q^2_{(cum)}$ (0.987) with a 2 PC model.

Although PCA is a very good method to depict the underlying variation between samples, there are several limitations associated with this model, one being that separation by PCA is achieved if maximum variation exists (Dao *et al.*, 2009). For the same and other maybe yet unidentified reasons, specifically pertaining to other biological systems, many reports on metabolomic analyses also include other models in parallel where multivariate statistics are used (Dao *et al.*, 2009; Maree and Viljoen, 2011). Here, OPLS-DA was used as a complimentary model to PCA. In the OPLS-DA score plot of control *vs*. 18 h treatment (Fig. 5A), good separation between samples originating from different treatment time intervals was observed. Similar to the loadings plot for PCA, the loadings S-plot for OPLS-DA (Fig. 5B) was used to identify possible biomarkers which are responsible for the separation observed on the score plot (Table 1).

Table 1: List of biomarkers of which the levels were found to be altered (up-regulated) due to INAP treatment in tobacco cell suspensions.

ID No.	Rt (min)	*m/z*	Metabolite name	Core structure
1	10.13	453.1065	1,2,4-Benzenetriol; 2-Me ether, 1-*O*-[3,4,5-trihydroxybenzoyl-(->6)-β-D-glucopyranoside]	Benzoic (gallic) acid
2	12.51	371.0947	Quinic acid; (-)-form, 4-*O*-(4-Hydroxy-3,5-dimethoxybenzoyl)	Benzoic (syringic) acid
3	9.30	329.0827	3,4-Dihydroxybenzoic acid; 3-Me ether, 4-*O*-β-D-glucopyranoside;	Benzoic (vanillic) acid
4	8.86	491.1425	3,4-Dihydroxybenzoic acid; 3-Me ether, 4-*O*-β-D-glucopyranoside, β-D-glucopyranosyl ester.	Benzoic (vanillic) acid
5	10.97	353.0861	3-*O* Caffeoylquinic acid	Cinnamic (caffeic) acid
6	8.92	445.136	4-O-beta-D-glucosyl-sinapate	Cinnamic (sinapic) acid
7	12.49	815.223	Kaempferol 3-rhamnosyl-(1->2)-galactoside-7-glucoside	Flavonoid
8	9.63	621.147	Kaempferol 3-[2'''-acetyl-alpha-L-arabinopyranosyl-(1->6)-galactoside]	Flavonoid

With the aid of open access online databases such as Chemspider and Dictionary of Natural Product (DNP), the identity (putative) of metabolites of which the levels were affected by INAP (Fig. 6, Table 1) were revealed. The distributions of these metabolites across different

treatment time intervals are shown in Fig. 7. Using n-Anova it was also shown that the identified metabolites differ significantly (p<0.01) in comparison to the control.

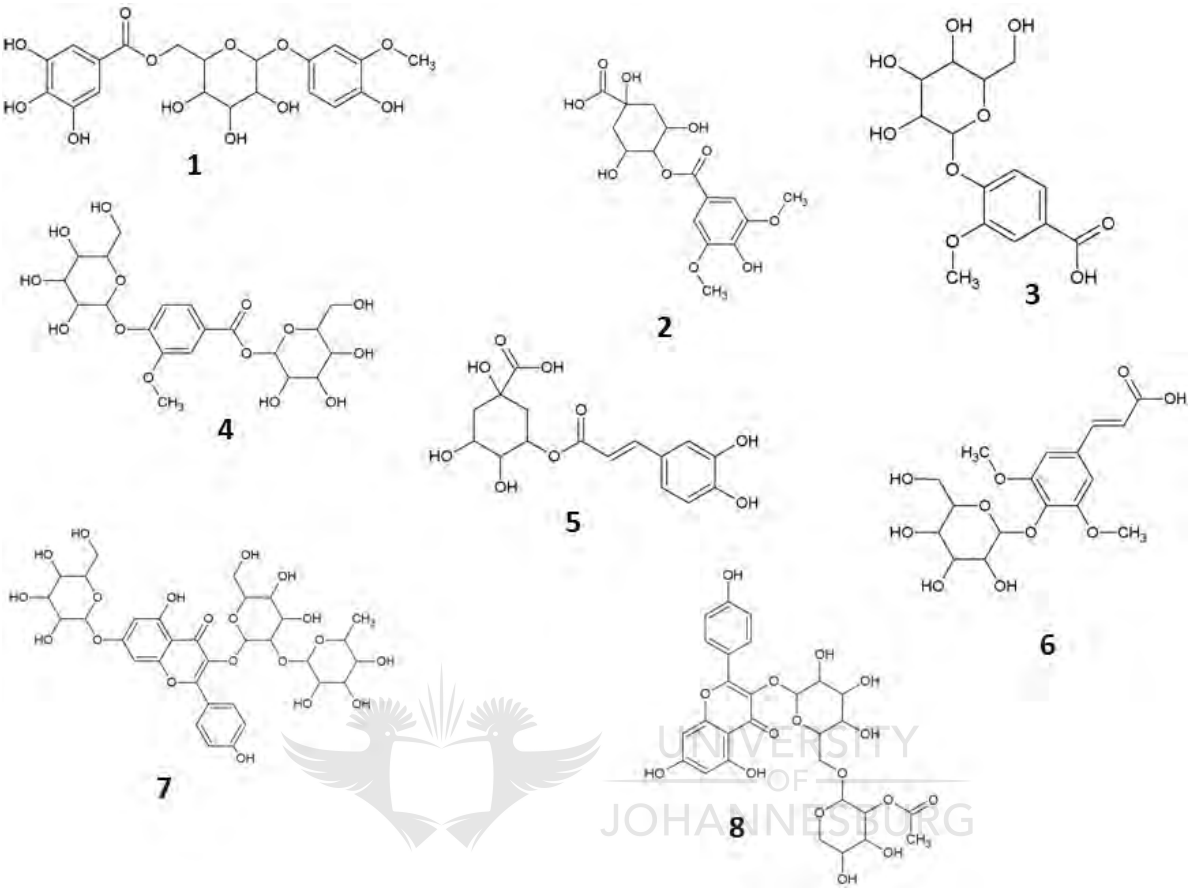


Fig. 6. Identified metabolites with the aid of PCA and OPLS-DA loading S-plots from the UHPLC-MS data analyses of INAP treated tobacco cell suspensions. From the DNP databases markers were identified as; 1. 1,2,4-Benzenetriol; 2-Me ether, 1-O-(3,4,5-trihydroxybenzoyl-(->6)-β-D-glucopyranoside); **2**. Quinic acid; (-)-form, 4-O-(4-Hydroxy-3,5-dimethoxybenzoyl); **3**. 3,4-Dihydroxybenzoic acid; 3-Me ether, 4-O-β-D-glucopyranoside; 4. 3-O Caffeoylquinic acid; **5**. 3,4-Dihydroxybenzoic acid; 3-Me ether, 4-O-β-D-glucopyranoside, β-D-glucopyranosyl ester; **6**. 4-O-beta-D-glucosyl-sinapate; **7**. Kaempferol 3-rhamnosyl-(1->2)-galactoside-7-glucoside and **8**. Kaempferol 3-[2'''-acetyl-alpha-L-arabinopyranosyl-(1->6)-galactoside].

From the structures of these identified molecules, it can be seen that (i) these are derived from shikimate - and phenylpropanoid pathway products, (ii) the cores resemble benzoic acid derivatives (**1**, **2**, **3**, **4**) or cinnamic acid derivatives (**5, 6**), and (iii) are either glycosylated (**1**, **3**,

4, 6) or quinilated (**2**, **5**). In addition, two different glycosylated derivatives of kaempferol, originating from the phenylpropanoid-flavonoid pathway, were also identified (**7, 8**).

From the above list of signatory biomarkers, it is clear that INAP treatment leads to major changes in phenylpropanoid composition. In tobacco cells, INAP is recognized by the enzymatic machinery of the phenylpropanoid pathway and feed into the cinnamic acid pathway. Novel enzyme-substrate combinations *in vivo* can lead to the biosynthesis of new, natural product-derived compounds (Pollier *et al*., 2011), and INAP is bioconverted to a molecule with a substitution pattern like ferulic acid (Madala *et al*., 2012). Some of the biotransformation events thus far identified include chemical modifications in the core structure. These include hydroxylation on the meta- and para- positions as well as subsequent methoxylation and glycosylation to result in 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone (Madala *et al*., 2012).

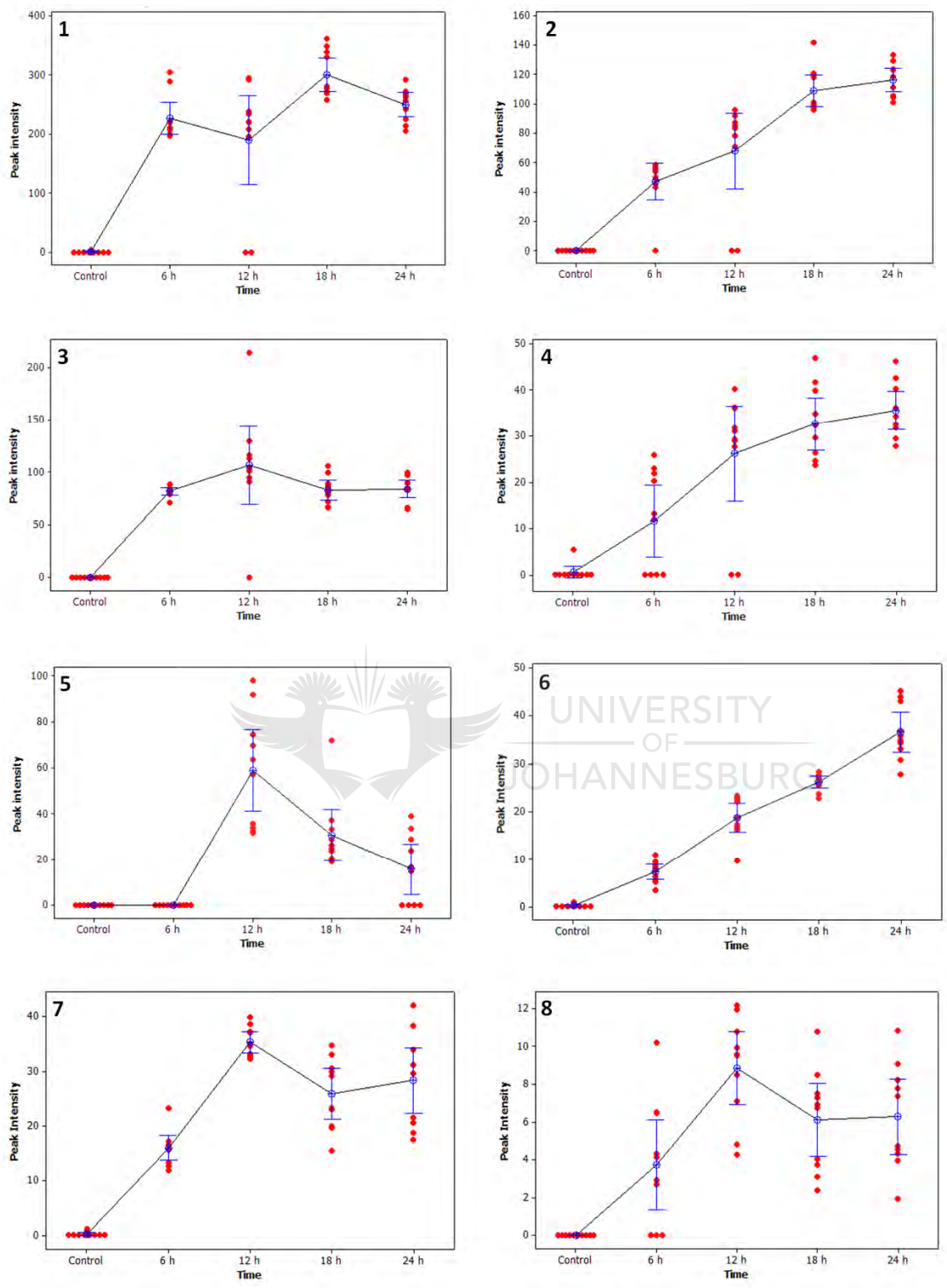


Fig. 7. Distribution graphs showing relative changes of identified metabolites across different treatment times. The numbers 1-8 represents molecules respectively, as shown in Fig. 6. The

target represents the mean and the bars represent the region of 95% confident interval of the mean. All the ions were found to differ significantly across different treatment times ($p<0.01$).

Cinnamic acid is synthesized from phenylalanine in the first step of phenylpropanoid biosynthesis (Dubery and Schabort, 1986, 1988). Phenylpropanoids are found throughout the plant kingdom, where they serve as essential protective and defensive components in the interaction of plants and the environment (Ververidis *et al*., 2007). In addition to phytoalexins and chemical defenses, cinnamate derivatives can also act as precursors of monolignols (Whetten and Sederoff, 1995), monomers that are polymerized to generate various forms of lignin and suberin, and are used as structural components of plant cell walls. Strenghtening of cell walls is part of the dynamic plant defense repertoire, and is known to restrict pathogen access to plant cells (Dubery and Smit, 1997; De Ascensao and Dubery, 2003). Derivatives identified in this study include caffeic acid (**5**) and sinapic acid (**6**) core structures.

Other molecules which were also found to be predominately induced by INAP are those belonging to the benzoic acid class. Plant benzoic acids and the respective derivatives are common and widespread mediators of plant responses to biotic and abiotic stress (Wildermuth, 2006), and can be synthesised *via* the shikimate/chorismate pathway or *via* cinnamic acid. Free and conjugated benzoic acid in tobacco plants and cell cultures are known to be induced upon elicitation of defense responses and may have a functional role as precursors of salicylic acid, which play an important role in activation of the plant 'defensome' (Chong *et al*., 2001). Identified metabolites harbor gallic acid (compound **1**), syringic acid (compound **2**) and vanillic acid (compounds **3** and **4**) core structures. The latter is known to be an antioxidant (Natella *et al*., 1999) and has antimicrobial activity (Maddox *et al*., 2010).

Molecules such as compound **5**, which was tentatively identified as O-caffeoylquinic acid (chlorogenic acid), form part of a large class of phenolic compounds that are formed as esters between quinic acid and one residue of trans-cinnamic acid derivatives, the most common ones being caffeic, *p*-coumaric- and ferulic acid (Clifford, 2000; Clifford, 2003; Clifford *et al*., 2006). In addition to playing a role in plant development (Franklin and Dias, 2011), chlorogenic acid is also believed to play a role during plant stress responses such as during plant : herbivore interaction (Jansen *et al*., 2009) and during abiotic stress responses (Torras-Cleveria *et al*., 2012).

Although the action mechanism of INAP is not known, its induction of these phenylpropanoid compounds warrants further investigation since phenylpropanoids are known to be triggered by different stressors such as pathogen attack, UV-irradiation, high light, wounding, nutrient deficiencies, temperature and herbicide treatment (Dixon and Paiva, 1995).

In conclusion, the use of multivariate statistical models such PCA and OPLS-DA to analyse UHPLC-MS data proved to be a powerful approach to study metabolic perturbations of cell suspensions treated with chemical inducers. The accumulation of all the metabolites shown in Fig. 6 is evidence that INAP induces metabolic changes in tobacco cell suspensions and these are known to have diverse functions in plants. Although the mode of action of INAP cannot be deduced from this data, this compound can be thought of as a chemical inducer of defence responses which may be recognised by enzymes involved in secondary plant metabolism and subsequently biotransformed to a stable molecule which resembles those of the phenylpropanoid pathway (Madala *et al*., 2012). The accumulation of these phenylpropanoid-derived compounds in general warrants further investigation. In addition, the current findings suggest that INAP can be used as a "priming" agent for defense responses in plants, thus enhancing natural defenses and contributing to protection of plants against pathogen attack.

5.5. References

Bednarek, P., Osbourn, A., 2009. Plant-microbe interactions: chemical diversity in plant defense. Science 324, 746–748.

Bell, A.A., 1981. Biochemical mechanisms of disease resistance. Ann. Rev. Plant Physiol. 32, 21–81.

Brown, M., Wedge, D.C., Goodacre, R., Kell, D.B., Baker, P.N., Kenny, L.C., Mamas, M.A., Neyses, L., Dunn, W.B., 2011. Systems biology automated workflows for accurate mass based putative metabolite identification in LC / MS-derived metabolomic datasets. Bioinformatics 27, 1108–1112.

Chong, J., Pierrel, M-A., Atanassova, R., Werck-Reichhart, D., Fritig, B., Saindrenan, P., 2001. Free and conjugated benzoic acid in tobacco plants and cell cultures. Induced accumulation upon elicitation of defense responses and role as salicylic acid precursors. Plant Physiol. 125, 318–328.

Clifford, M.N., 2000. Chlorogenic acids and other cinnamates – nature, occurrence, dietary burden, absorption and metabolism. J. Sci. Food Agr. 80, 1033–1042.

Clifford, M.N., 2003. The analysis and characterization of chlorogenic acids and other cinnamates. In C. Santos-Buelga & G. Williamson (Eds.), Methods in polyphenol analysis (pp. 314–337). Cambridge: Royal Society of Chemistry.

Clifford, M.N., Knight, S., Surucu, B., Kuhnert, N., 2006. Characterization by LC–MS^n of four new classes of chlorogenic acids in green coffee beans: Dimethoxycinnamoylquinic acids, diferuloylquinic acids, caffeoyl dimethoxy-cinnamoylquinic acids, and feruloyl dimethoxycinnamoylquinic acids. J. Sci. Food Agr. 54, 1957–1969.

Cloarec, O., Dumas, M.E., Trygg, J., Craig, A., Barton, R.H., Lindon, J.C., Nicholson, J.K., Holmes, E., 2005. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. Anal. Chem. 77, 517–526.

Dao, T.T.H., Puig, R.C., Kim, H.K., Erkelens, C., Lefeber, A.W.M., Linthorst, H.J.M., Choi, YH., Verpoorte R., 2009. Effect of benzothiadiazole on the metabolome of *Arabidopsis thaliana*. Plant Physiol. Biochem. 47, 146–152.

De Ascensao, A.R.D.C.F., Dubery, I.A., 2000. Panama disease: Cell wall reinforcement in banana roots in response to elicitors from *Fusarium oxysporum* f. sp. *cubense* race four. Phytopathology 90, 1173–1180.

De Ascensao, A.R.F.D.C., Dubery, I.A., 2003. Soluble and wall-bound phenolics and phenolic polymers in *Musa acuminata* roots exposed to elicitors from *Fusarium oxysporum* f.sp. *cubense*. Phytochemistry 63, 679–686.

Dixon, R.A., Harrison, M.J., Lamb, C.J., 1994. Early events in the activation of plant defense responses. Ann. Rev. Phytopathol. 32, 479–501.

Dixon, R.A., Paiva, N.L., 1995. Stress-induced phenyl-propanoid metabolism. Plant Cell 7, 1085–1097.

Dubery, I.A., Louw, A.E., Van Heerden, F.R., 1999. Synthesis and evaluation of 4-(3-methyl-2 butenoxy) isonitrosoacetophenone, a radiation-induced stress metabolite in Citrus. Phytochemistry 50, 983–989.

Dubery, I.A., Smit, F., 1997. Cell wall reinforcement in cotton hypocotyls in response to a *Verticillium dahliae* elicitor. Phytochemistry 44, 811–815.

Dubery, I.A., Schabort, J.C., 1986. Phenylalanine ammonia-lyase from *Citrus sinensis*: purification by hydrophobic interaction chromatography and physical characterization. Biochem. Int. 13, 579–589.

Dubery, I.A., Schabort, J.C., 1988. Kinetic and regulatory properties of phenylalanine ammonia lyase from *Citrus sinensis*. Int. J. Biochem. 20, 217–222.

Esterhuizen, L.L., Meyer, R, Dubery, I.A. 2006. Antioxidant activity of metabolites from *Coleonema album* (Rutaceae). Nat. Prod. Commun. 1, 367–375.

Fiehn, O., 2002. Metabolomics–the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

Franklin, G., Dias, A.C.P., 2011. Chlorogenic acid participates in the regulation of shoot, root and root hair development in *Hypericum perforatum*. Plant Physiol. Biochem. 49, 835–842.

Gatz, C., 1997. Chemical control of gene expression. Ann. Rev. Plant Physiol. Plant Mol. Biol. 48, 89–108.

Gatz, C., Lenk, I., 1998. Promoters that respond to chemical inducers. Trends Plant Sci. 3, 352–358.

Görlach, J., Volrath, S., Knauf-Beiter, G., Hengy, G., Beckhove, U., Kogel, K-H., Oostendorp M., Staub, T., Ward, E., Kessmann, H., Ryals, J., 1996. Benzothiadiazole, a novel class of inducers of systemic acquired resistance, activates gene expression and disease resistance in wheat. Plant Cell 8, 629–643.

Greenberg, J.T., 1997. Programmed cell death in plant-pathogen intercations. Ann. Rev. Plant Physiol. Plant Mol. Biol. 48, 525–545.

Hammond-Kosack, K.E., Jones, J.D.G., 1996. Resistance gene-dependant plant defense responses. Plant Cell 8, 1773–1791.

Jakab, G., Cottier, V., Toquin, V., Rigoli, G., Zimmerli, L., Métraux, J-P., Mauch-Mani, B., 2001. β-aminobutyric acid-induced resistance in plants. Eur. J. Plant Pathol. 107, 29–37.

Jansen, J.J., Allwood, J.W., Marsden-Edwards, E., van der Putten, W.H., Goodacre, R., van Dam, N.M., 2009. Metabolomic analysis of the interaction between plants and herbivores. Metabolomics 5, 150–161..

Jolliffe, I.T., 2002. Principal Component Analysis. 2nd ed. Springer, New York.

Jung, H.W., Tschaplinski, T.J., Wang, L., Glazebrook, J., Greenberg, JT., 2009. Priming in systemic plant immunity. Science 324, 89–91.

Liu, N.Q., Cao, M., Frédérich, M., Choi, Y-H., Verpoorte, R., Van der Kooy, F., 2010a. Metabolomic investigation of the ethnopharmacological use of *Artemisia afra* with NMR spectroscopy and multivariate data analysis. J. Ethnopharmacol. 128, 230–235.

Liu, F., Wei, F., Wang, L., Liu, H., Zhu, X., Liang, Y., 2010b. Riboflavin activates defense responses in tobacco and induces resistance against *Phytophthora parasitica* and *Ralstonia solanacearum*. Physiol. Mol. Plant Pathol. 74, 330–333.

Madala, N.E., Steenkamp, P.A., Piater, L.A., Dubery, I.A., 2012. Biotransformation of isonitrosoacetophenone (2-keto-2-phenyl-acetaldoxime) in tobacco cell suspensions. Biotechnol. Lett. 37, 1351–1356.

Maddox, C.E., Laur, L.M., Tian, L., 2010. Antibacterial activity of phenolic compounds against the phytopathogen *Xylella fastidiosa*. Curr. Microbiol. 60, 53–58.

Mahadevan, S., 1973. Role of oximes in nitrogen metabolism in plants. Annu. Rev. Plant. Physiol. 24, 69–88.

Maree, J.E., Viljoen, A.M., 2011. Fourier transform near- and mid-infrared spectroscopy can distinguish between the commercially important *Pelargonium sidoides* and its close

taxonomically *P. reniforme*. Vib. Spectrosc. 55, 146–152.

McDowell, J.M., Dangl, J.L., 2000. Signal transduction in the plant immune response. Trends Biochem. Sci. 25, 79–82.

Métraux, J-P., Ahl Goy, P., Staub, T., Speich, J., Steinemann, A., Ryals, J., Ward, E., 1991. Induced systemic resistance in cucumber in response to 2,6-dichloro-isonicotinic acid and pathogens. Pages 432–439 in: Advances in molecular genetics of plant-microbe interactions, vol. 1. H. Hennecke and D. P. S.Verma, eds. Kluwer, Dordrecht, The Netherlands.

Moller, B.L., 2010. Dynamic metabolons. Science 330, 1328–1329.

Natella, F., Nardini, M., Di Felice, M., Scaccini, C., 1999. Benzoic and cinnamic acid derivatives as antioxidants: structure-activity relation. J. Agric. Food Chem. 47, 1453–1459.

Odjakova, M., Hadjiivanova, C., 2001. The complexity of pathogen defense in plants. Bulgarian J. Plant Physiol. 27, 101–109.

Oostendorp, M., Kunz, W., Dietrich, B., Staub, T., 2001. Induced resistance in plants by chemicals. Eur. J. Plant Pathol. 107, 19–28.

Pollier, J., Mosesab, T., Goossens, A., 2011. Combinatorial biosynthesis in plants: A review on its potential and future exploitation. Nat. Prod. Rep. 28, 1897–1916.

Ryals, J.A., Neuenschwander, U.H., Willits, M.G., Molina, A., Steiner, H., Hunt, M.D., 1996. Systemic acquired resistance. Plant Cell 8, 1809–1819.

Sanabria, N.M., Dubery, I.A., 2006. Differential displayprofiling of the *Nicotiana* response to LPS reveals elements of plant basal resistance. Biochem. Biophy. Res. Comm. 344, 1001–1007.

Torras-Claveria, L., Jáuregui, O., Codina, C., Tiburcio, A.F., Bastida, J., Viladomat, F., 2012. Analysis of phenolic compounds by high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry in senescent and water-stressed tobacco. Plant Sci. 182, 71–78.

Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). J. Chemometr. 16, 119–128.

Van Loon, L.C., 1997. Induced resistance in plants and the role of pathogenesis-related proteins. Eur. J. Plant Pathol. 103, 753–765.

Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., Panopoulos, N., 2007.

Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part II: Reconstruction of multienzyme pathways in plants and microbes. Biotechnol. J. 2, 1235–1249.

Whetten, R., Sederoff, R., 1995. Lignin biosynthesis. Plant Cell 7, 2033–2046.

Wildermuth, M.C., 2006. Variations on a theme: synthesis and modification of plant benzoic acids. Curr. Opin. Plant Biol. 9, 288–296.

Zwiegelaar, M., Dubery, I.A., 2006. Early activation of cell wall strengthening-related gene transcription in cotton by a *Verticillium dahliae* elicitor. SA. J. Bot. 72, 467–472.

5.6. Supplementary files

Electronic supplementary file S1 and S2: List of metabolites identified with the aid PUTMEDID_LC-MS workflow using both negative and positive ionization UHPLC-MS data respectively, see CD attached.

Chapter 6: Metabolic perturbations induced by isonitrosoacetophenone in *Arabidopsis thaliana* plants results in an enhanced defensive environment.

6.1. Abstract

During the life-cycle, plants are constantly exposed to numerous biotic or abiotic stress factors. Pathogens and pathogen-derived molecules are the most studied inducers of plant defense responses due to the devastating economical consequences. However, synthetic and naturally occurring molecules have also been used to induce various types of resistance in plants against pathogens. Such molecules include those derived from pathogens but recently, plant-derived molecules have also been included. In the current study, a plant-derived oxime molecule, 2-isonitrosoacetophenone (INAP), was used to trigger metabolic changes as monitored in the metabolome of treated plants. UHPLC-MS in conjunction with principal component analysis (PCA) and orthogonal projection to latent structures discriminant analysis (OPLS-DA) was applied to investigate the metabolic changes in *Arabidopsis thaliana* plants treated with INAP. These chemometric methods managed to reveal metabolites found to be significantly present as a response to the treatment. These include bio-transformation events (phenylacetaldoxime-glucoside and mandelonitrile-glucoside) as well as those of which the levels are affected by the treatment (benzoic acid, benzoylglucoside and scopoletin). Using *in planta* bacterial growth evaluations, INAP treatment was furthermore found to induce an anti-microbial environment *in vivo*.

***Keywords*:** Isonitrosoacetophenone, metabolomics, metabolites, PUTMEDID_LC-MS workflow, PCA, OPLS-DA, *Arabidopsis thaliana*, induced resistance, priming.

6.2. Introduction

Unlike mammals, plants do not possess adaptive immunity and solely rely on innate immunity to fight against the wide spectrum of stress factors that are constantly encountered. The term induced resistance (IR) is commonly used to describe this type of plant innate immunity and can be defined as a physiological state of enhanced defensive capacity, elicited by specific stimuli, and is potentiated against subsequent biotic challenges (van Loon *et al*., 1998). Diverse forms of IR exist and are mainly differentiated on the basis of implementation and outcomes (Pieterse *et al*., 2009).

Systemic acquired resistance (SAR), which was previously known as physiological acquired immunity, is the common form of IR. By definition SAR is a physiological response in which plants exhibit a long-lasting resistance due to infection, and results in a subsequent stronger response towards successive infection as well as reduced disease symptoms (Ryals *et al*., 1996). Another similar type of IR exists and is commonly referred to as induced systemic resistance (ISR). Although the term 'induced' is absent in SAR and biased towards ISR, it is worth noting that both are manifested upon induction of plant defense. These two forms of defense are quite similar, but can be differentiated physiologically rather than phenotypically since both are activated differently and independently. SAR is triggered by biotrophic organisms and is associated with the accumulation of salicylic acid (SA) (Ryals *et al*., 1994; Zimmerli *et al*., 2000, 2001). In contrast, beneficial microorganisms known as plant growth-promoting rhizobacteria (PGPR) are known to induce ISR (Ramamoorthy *et al*., 2001). In the latter case, the phyto-hormones jasmonic acid and ethylene (JA / ET), generated during attack by necroptrohic pathogens, are believed to play a central role (Thomma *et al*., 1998; Dubery *et al*., 2000; Glazebrook, 2005). A third form of IR, known as priming, has also been proposed but, unlike the former two, this type is not fully characterized. Priming implies to prepare or to make ready, and in the context of plant defense responses, this is a physiological process by which a plant is predisposed to respond more rapidly and strongly to future biotic or abiotic stresses (Conrath *et al*., 2006; Goellner and Conrath, 2008). The condition of readiness achieved by priming has been termed the 'primed state' (Conrath *et al*., 2006) but, unlike SAR and ISR, does not result in full implementation of plant defense responses and, as such, does not impair the host plants' energy pool.

In most cases, IR is commonly thought to be exclusively triggered by pathogens, however, it should kept in mind that most studies aiming to understand IR have been conducted using live pathogens. It is only recently that several synthetic and naturally occurring molecules have been used to induce different forms of IR, for example chemicals such as isonicotinic acid (INA) (Gorlach *et al.,* 1996), β-aminobutyric acid (BABA) (Ton *et al*., 2005), sucrose (Gomez-Ariza *et al*., 2007), riboflavin (Zhang *et al*., 2009), saccharin (Walters *et al*., 2009), hexanoic acid (Vicedo *et al*., 2009), and azealic acid (Jung *et al*., 2009). The biological mechanisms of action of these chemicals are not fully understood and more studies need to be conducted in an attempt to answer this question. An appropriate example is benzothiadiazole [benzo-(1,2,3)-thiadiazole-7-carbothioc acid S-methyl ester] (BTH), a functional analogue of SA and key player in SAR (Gorlach *et al*., 1996). It is one of the most researched molecules in the context of plant defense studies based on its ability to induce a more effective plant resistance response. This molecule has also been shown to prime the plant defense machinery against different pathogens (Friedrich *et al*., 1996; Gorlach *et al*., 1996).

The implementation of IR is not readily notable by mere visual inspection of infected plants and, as such, scientists aim to understand how this complicated response is initiated. Several biological approaches have been utilized to achieve the above and all differ with the mode of analysis and detection. IR can, for instance, be investigated using transcriptomics, and marker genes encoding proteins such as NPR1 (non-expressor of PR gene 1) (Conrath *et al*., 2002), mitogen activated protein kinase (MAPK)-3 and -6 (Beckers *et al*., 2009), flavin-dependant monooxygenase (FMO1) (Mishina and Zeier, 2007), pathogenesis-related (PR) proteins (Lawton *et al*., 1996; Zhang *et al*., 2009) and the lipid transfer proteins (LTPs) (Maldonado *et al*., 2002; Jung *et al*., 2009) have been suggested. It is also interesting to note that several promoters for genes encoding proteins functioning towards plant defense responses, such as that of PR-1, are known to respond to several chemical activators of defense (especially BTH) (Gatz and Lenk, 1998; Padidam, 2003).

Using bacterial lipopolysaccharide (LPS) as an inducer of SAR (Coventry and Dubery, 2001) we also recently showed that both gene expression studies (Sanabria and Dubery, 2006; Madala *et al.,* 2011, 2012) and proteomics (Piater *et al*., 2004; Gerber *et al*., 2006, 2008) contribute to a comprehensive understanding of activation of plant defense responses. Recently, it was shown that metabolites can also be used in this regard (Beets *et al*., 2012, Tugizimana *et*

al., 2012). By virtue of being the end product of most cellular metabolism, metabolites act as regulatory components of biological information flow and hence the accumulation and/or changes allow a complete snapshot of the physiological status of the cell (Fernie *et al*., 2004; Ryan and Robards, 2006; Seger and Sturm, 2007), referred to as metabolomics (Nicholson *et al*., 1998; Oliver *et al*., 1998).

In the current study, isonitrosoacetophenone (INAP, or 2-keto-2-phenyl-acetaldoxime), a structural analogue of an oxidative stress metabolite with anti-oxidant and anti-fungal activity, 4-(3-methyl-2-butenoxy)-isonitrosoacetophenone (Dubery *et al.* 1999), was used to trigger metabolic changes in *Arabidopsis thaliana* plants to further our existing knowledge on oxime metabolism in plants. Oximes are rare functional groups in natural products but they occur in a variety of phyla, e.g., sponges, bacteria, fungi, and plants (Almeida *et al.,* 2011). In plants, oximes participate in different biosynthetic pathways and have been shown to act as intermediates of a select few metabolic pathways subject to controls that result in variation in both the type and amount of end product formed (Mahadevan, 1973). The most notable involvement of oximes in plants is during the biosynthesis of stress-related compounds known as glucosinolates and cyanogenic glycoside which function against pest (herbivore) attack (Moller, 2010).

Untargeted UHPLC-MS-based metabolic analysis in combination with PCA and OPLS-DA was used to investigate the metabolic-inducing (or IR) ability of INAP and to explain the existence of oximes in non-cyanogenic plants. Bacterial growth in INAP pre-treated plants was further used to understand the IR-inducing ability of this oxime.

6.3. Material and methods

6.3.1. Plant treatment and metabolite extraction

Thirty (30) day old *Arabidopsis thaliana* (Columbia) plants were treated by means of foliar spraying with a solution of INAP (1 mM) in 10 mM $MgCl_2$, while control plants were only sprayed with 10 mM $MgCl_2$. Plants were allowed to incubate for a period of 18 h at 25ºC prior to metabolite extraction. Following the treatment period, metabolites were extracted from leaves (2 g) with 100% methanol (20 mL) using an ultraturax rotating blade homogenizer. To aid maximum extraction, the homogenates were further sonicated for 20 min and heated at 60ºC for

10 min. The homogenates were then centrifuged at 10 000 x *g* for 10 min, and the supernatants transferred to a round bottom flask prior to reducing the volume to approximately 2 mL on a rotary evaporator. The resulting volume was further dried to completeness using a speed vacuum centrifuge (R.C 10.09; Jouan, France) operating with constant heating at 50ºC. The precipitates were re-dissolved in 400 µL 50% (v/v) methanol in water and filtered through a 0.22 µm filter using a 1 mL sterile syringe. Unless stated elsewhere, all reagents and solvents were of UHPLC grade.

6.3.2. UHPLC-HD-ESI-MS analysis

Five (5) µL methanol extract was analyzed on the Waters Acquity UHPLC-high definition MS instrument equipped with the Acquity CSH C18 column (150 mm × 2.1 mm with particle size of 1.7 µm) (Waters Corporation, Milford, USA). The composition of mobile phase A consisted of 0.1% formic acid in deionized water and mobile phase B consisted of 0.1% formic acid in methanol. The length of chromatographic separation was 20 min with the gradient conditions: 5% B over 0.1-1.0 min, 5-95% B over 1.0-16.0 min, held constant at 95% B over 16.0-17.0 min, 95-5% B over 17.0-18.0 min and final wash at 5% B over 18-20 min at constant flow rate of 0.4 mL/min. Chromatographic separation was monitored using both a photodiode array detector (PDA) (100-500 nm) and a mass spectrometer (MS) detector operating in positive and negative ionization mode. The MS conditions were as follows: capillary voltage of 2.5 kV, sample cone voltage of 17 V, multichannel plate detector voltage of 1750 V, source temperature of 120°C, desolvation temperature of 400°C, cone gas flow of 50 L/h and desolvation gas flow of 450 L/h. The mass spectrometric full scan data were acquired from 100-1000 Da with a scan time of 0.1 sec, interscan delay of 0.02 sec. Data was centroided and mass spectra corrected in real time by an external reference standard consisting of leucine-enkephalin (5 pg/mL) using a lockmass sprayer interface and a lockmass flow rate of 0.2 mL/min.

6.3.3. Data analysis

For multivariate data analyses, both negative and positive ionization raw data files were visualized by MassLynx XS^{TM} software version 4.1 (Waters Corporation, Milford, USA) and further analyzed by MarkerLynx^{TM} software. The analysis parameters for MarkerLynx^{TM} were set to analyse between 3-17 min, mass range 100-1000 Da, mass tolerance 0.01 Da, retention time (Rt) window 0.2 min, mass window of 0.02 Da and isotopic peaks were included for the analysis. The signal believed to be residual INAP was also removed for the subsequent analysis. The dataset obtained from MarkerLynx^{TM} processing was exported to the SIMCA-P software version 12.0 (Umetrics, Umea, Sweden) programme for PCA analysis. Unless stated otherwise, PCA and OPLS-DA models were centered and then Pareto scaled using SIMCA-P software. Sample classification based on the metabolite content was achieved by PCA and OPLS-DA score plots, and metabolites of which the levels were found to be affected by INAP treatment were highlighted using the PCA loadings plot and OPLS-DA based S-plot. From the S-plot, only signatory biomarkers (metabolites) with the correlation coefficient [P(corr)] of $\geq$ 0.8 and covariance coefficient (p1) $\geq$ 0.05, were selected and the respective *m/z* values utilized to calculate the elemental composition for metabolite identification.

Metabolite identification has proven to be the most difficult aspect of the metabolomics workflow in the past (Dunn *et al.,* 2012a) and, as such, for the current study a more compressive/advanced route was preferred. Here, Markerlynx XS^{TM} results were modified to a version compatible/similar to that required for the PUTMEDID_LC-MS workflow (Brown *et al.,* 2011). These workflows operate on the Taverna workbench (http://www.taverna.org.uk) and are comprised of three main functions which include (a) correlation analysis, (b) metabolic feature annotation, and (c) metabolite annotation. The basic functioning of this workflow lies in its ability to allow for peaks (accurate *m/z*) which share similar features, such as Rt, to be grouped together using basic correlation coefficients such as Pearson or Spearman algorithms. Further grouping of peaks takes place and once the peaks which belongs to a specific metabolite exist as "features" (isotope, adduct, dimer, others), the next step then involves calculation of the elemental composition/molecular formula of each accurate *m/z*. Lastly, calculated molecular formulae are then used to tentatively identify the metabolites using freely available online

databases such as DNP, Chemspider, AraCyC, PlantCYC, KEGG, LipidMAPS, KNApSAcK and METLIN. This can be done automatically if the database is downloadable, or manually using the search engine window. It is worth stating that such workflows are only applicable for data generated using an instrument which is capable of measuring high accurate mass (< 5 ppm) such as the Q-TOF-MS, which was used during the current study.

6.3.4. *In planta* growth evaluation

Arabidopsis plants (3 per condition) were sprayed with INAP (1 mM) or 10 mM $MgCl_2$ (control) on two consecutive days and on the third day, plants were inoculated with *Pseudomonas syringae* pv. *maculicola* (OD_{600} = 0.01) in 10 mM $MgCl_2$ by means of foliar spraying (control plants were again only sprayed with 10 mM $MgCl_2$). Three days post-inoculation (dpi), leaves were harvested, weighed, sterilized, rinsed and homogenized in 10 mM $MgCl_2$, and plant debris removed by centrifugation at 500 x *g* for 10 min. The resulting supernatant was diluted to the desired concentration and 100 µL spread on King's B medium plates and kept at 28ºC for two days. Following growth, colonies were manually counted and quantified as follows: cfu/g = [colony number x total volume (µL) x dilution factor] / [coating volume (µL) x tissue weight (g)]. The resulting data was used to construct the bar graph and also for statistical significant test (Student t-test) using Microsoft Excel.

6.4. Results and discussion

INAP (or 2-keto-2-phenyl-acetaldoxime) is an oxime molecule with structural similarities to citaldoxime (4-(3-methyl-2-butenoxy)-isonitrosoacetophenone), a novel metabolite found to accumulate in citrus peel as a result of oxidative stress due to gamma radiation treatment (Dubery *et al*., 1999). In the current study, the effect of INAP on the metabolome of *A. thaliana* was investigated.

Metabolomics is the latest and fastest growing amongst the '–omics' field of studies and, by definition, entails the comprehensive analysis of metabolites in a biological sample at any given physiological status (Fiehn, 2002; Sumner *et al*., 2003). By virtue of being the end product of most cellular metabolism, metabolites act as regulatory components thereof and hence the accumulation allows a complete snapshot of the physiological status of the cell (Fernie *et al*.,

2004; Ryan and Robards, 2006; Seger and Sturm, 2007). One notable characteristic of plants is that they are capable of producing a vast number of natural compounds with different functionalities. These are sometimes referred to as secondary metabolites, and play a role in the basic developmental cycles of plants and also during interactions with the environment (biotic or abiotic) (Kutchan, 2001).

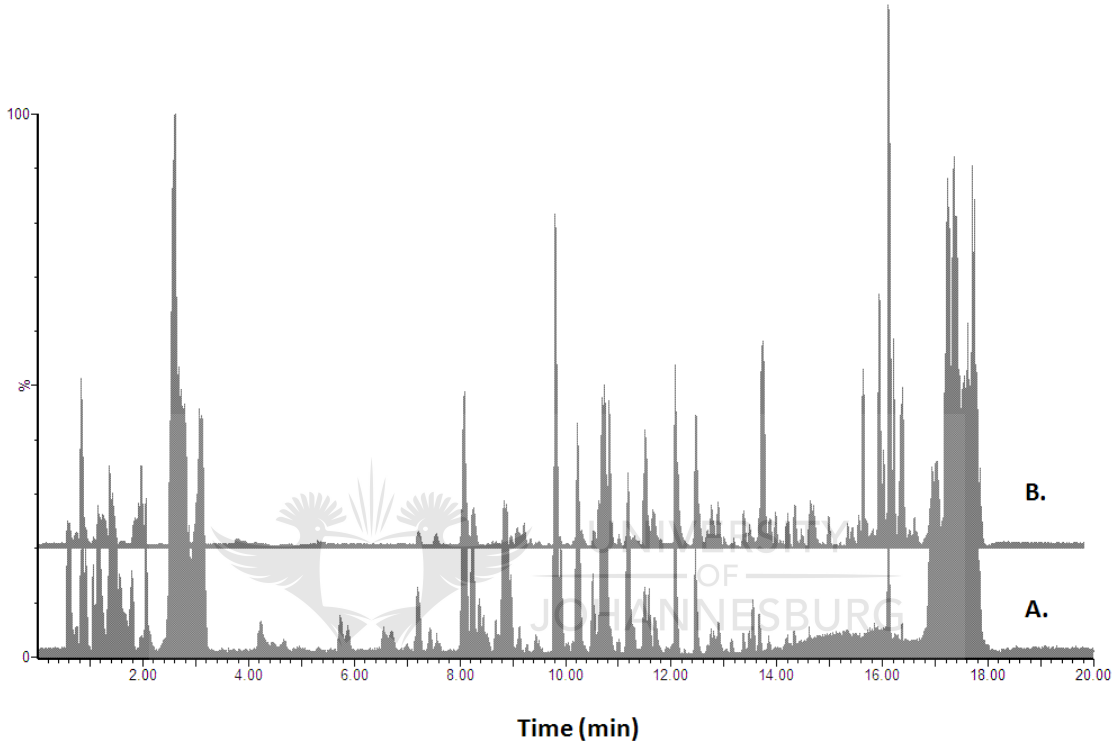


Figure 1: Representative UHPLC-MS chromatograms showing qualitative separation of all detected biomarkers in methanolic extracts from *A. thaliana* leaves treated for 18 h with INAP under negative (A) and positive (B) ionization conditions.

In order to establish the metabolic fate of INAP in *A. thaliana* plants, metabolic profiling of treated plants was performed and compared to untreated counterparts using UHPLC-MS as the preferred analytical platform, in combination of PCA and OPLS-DA analysis of the resulting data. Both ESI negative and positive ionization data clearly show that several metabolites were detected (Fig. 1A, B). PCA and OPLS-DA models allow for comprehensive data mining and subsequent identification of signatory metabolites (biomarkers) of which the levels are affected by the treatment under investigation. From the current study, INAP was shown to induce metabolic perturbations in *A. thaliana* plants as is evident when UHPLC-MS chromatograms of

treated and untreated samples are compared. PCA score plots of data obtained from both negative and positive ionization managed to separate control and treated samples into distinct groups (Fig. 2 A, B).

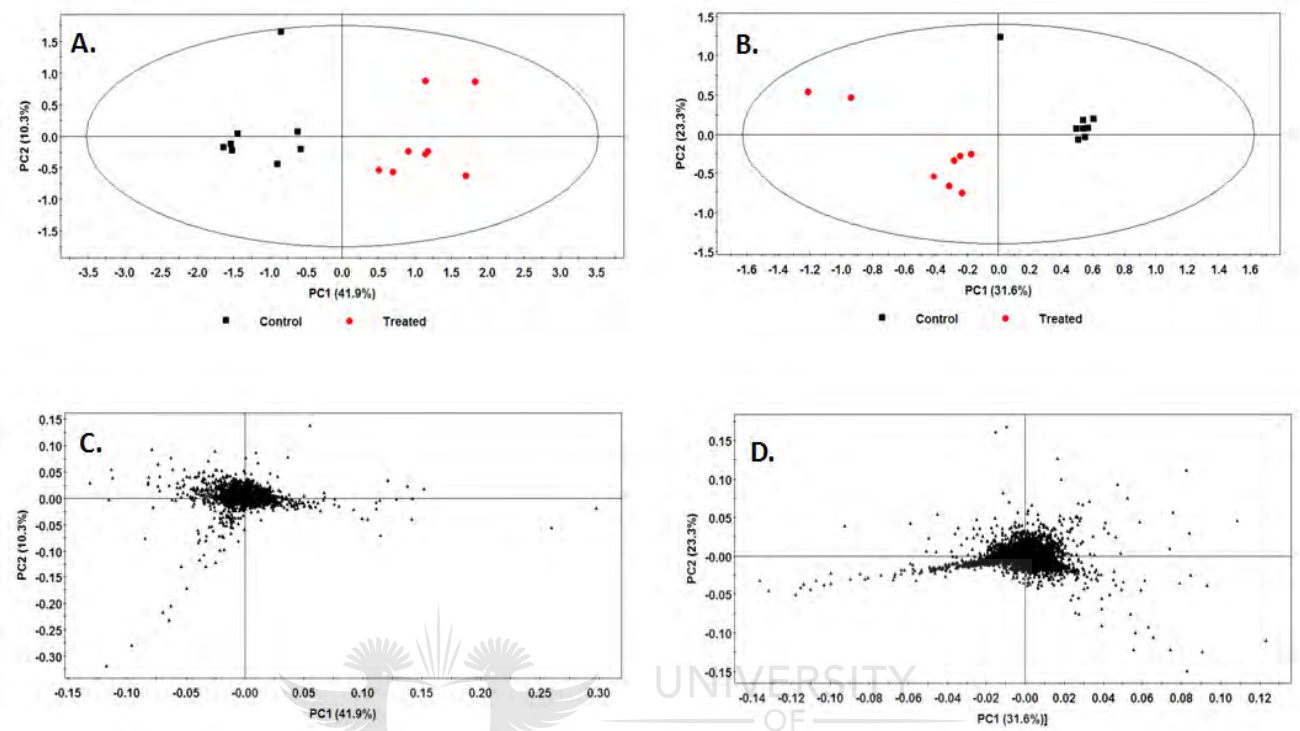


Figure 2: PCA score plots, based on the UHPLC-MS chromatograms from negative (A) and positive (B) ionization data, showing clustering of samples from *Arabidopsis* plants treated with INAP for 18 h and the corresponding loadings plot (C and D), showing biomarkers responsible for the clustering observed in A and B respectively. The ellipse represents Hotelling's T2 with 95% confidence.

PCA is an unsupervised model and can be mathematically defined as an orthogonal linear transformation of possibly correlated variables into a smaller number of uncorrelated variables called principal components, where the greatest variance within the data by any projection is explained on the first co-ordinate (called the first principal component, PC) and the least variance is explained/projected by subsequent principal components (Jolliffe, 2002). This reduction allows samples to be easily classified based on the metabolite content. Here, samples from different biological backgrounds are clearly separated into distinct clusters, and samples that group together can be referred to as a specific 'metabolic phenotype' (Fiehn *et al*., 2000). It

is thus clear that INAP induced metabolite perturbations in treated plants since such metabolites are clearly projected as outliers (away from the center) on the PCA loadings plots (Fig. 2 C, D) in the direction in which the samples are scattered on the PCA score plot. Thus, when the score and loadings plot are superimposed, only metabolites (on the loadings plot) which lie to the direction of either sample grouping on the score plot have influence on that particular group. Although PCA has been used intensively in the field of metabolomics, it has disadvantages of its own. It is important to note that separation by PCA is achieved if maximum variation exists (Dao *et al*., 2009), thus this type of analysis only evaluates global patterns (maximum variation) within the data and, as such, it is not a good tool for revealing local phenomena. Apart from this reason, there are more negative sentiments associated with PCA (Van der Greef and Smilde, 2005), and metabolomic investigations usually require alternative methods in conjunction with PCA in order to strengthen the conclusions. In the current study, OPLS-DA was chosen as an alternative model. Although OPLS-DA is a supervised model, the separation achieved is quite similar to that seen with PCA, however, information between and within classes is more comprehensively explained (Wiklund *et al*., 2008). From the current study, control and treated samples were found to form distinct clusters on the OPLS-DA score plot (Fig. 3 A, B) and such separation is more pronounced on OPLS-DA as compared to PCA. Using the S-loadings plot (Fig. 3 C, D) (Wiklund *et al*., 2008), metabolites of which the levels are significantly altered by INAP treatment can be easily identified. Such metabolites are projected on the extreme ends of the 'S'-like figure and, similarly to the loadings plot, biomarkers which are projected towards the direction of particular cluster on the score plot influence that particular group/cluster.

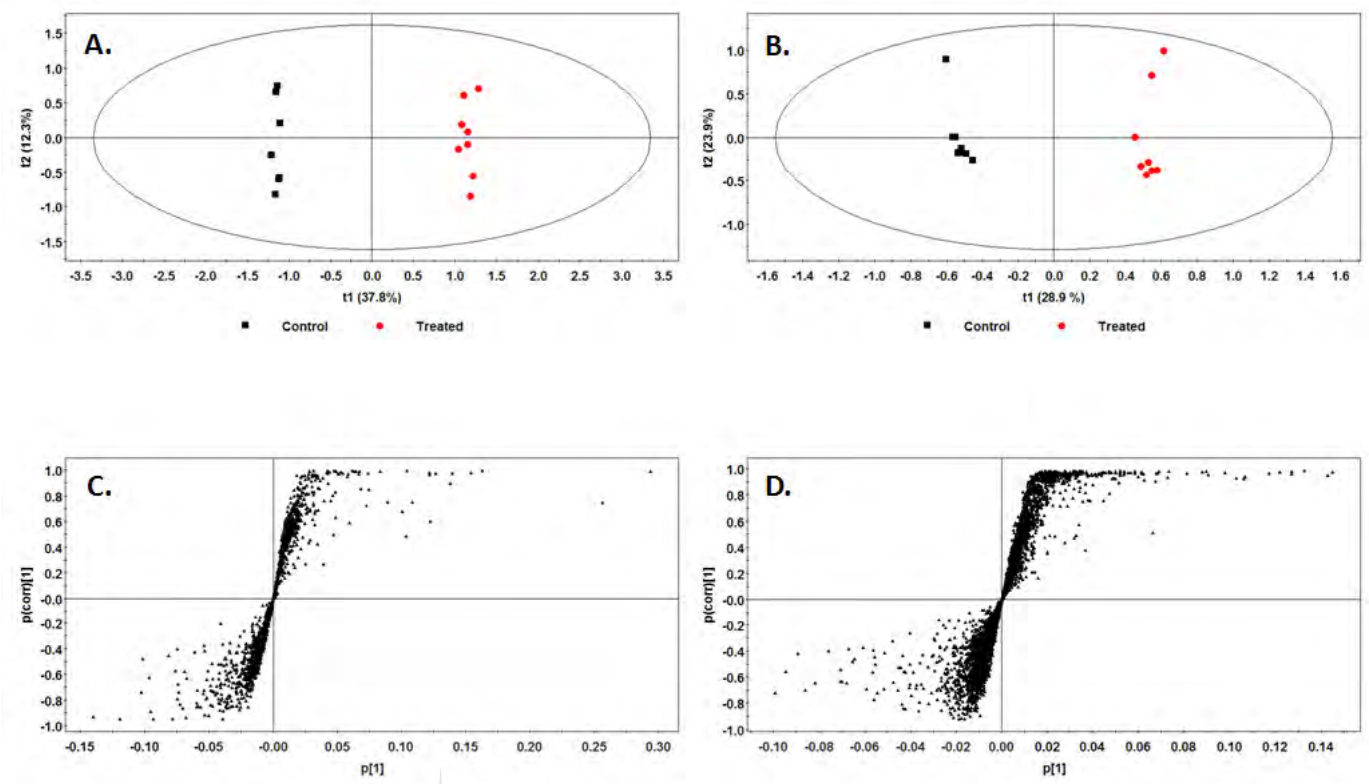


Figure 3: OPLS-DA score plots, based on the UHPLC-MS chromatograms from negative (A) and positive (B) ionization data, showing clustering of samples from *Arabidopsis* plants treated with INAP for 18 h and the corresponding loadings S-plot (C and D), showing biomarkers responsible for the clustering observed in A and B respectively. The ellipse represents Hotelling's T2 with 95% confidence.

The last step during metabolomic studies is the identification of all detected metabolites and, most importantly, those of which the levels are affected by the treatment in question. Contrary to GC-MS for which extensive metabolite libraries exist, there are very limited libraries for LC-MS (Von Roepenack-Lehaye *et al*., 2004), and thus metabolite identification in this regard presents a challenge. It is only with the recent introduction of Q-TOF instruments that LC-MS-based metabolomics data is becoming useful for metabolite identification. These instruments combine the accurate mass analysis ability of a TOF analyzer with that of ESI ionization mode which results in high sensitivity, high mass resolution and high mass accuracy (< 5 ppm) (Chernusheshevich *et al*., 2001). To date, Q-TOF instruments have been applied

extensively in the field of metabolite analysis (Plumb *et al*., 2005; Lu *et al*., 2008; Kim *et al*., 2011). Even though the application of Q-TOF is becoming the norm in metabolite profiling and related studies, there still exist challenges such as metabolite identification (Dunn *et al*., 2012a). Apart from the main aim of the current study, a novel way for metabolite identification was also investigated and proved promising with a high degree of confidence. This approach is based on the PUTMEDID_LC-MS workflow (Brown *et al*., 2011), and several reports have shown the ability of this workflow in the identification of metabolites (Dunn *et al*., 2012b). Most, however, have been performed on animal models. Traditionally, metabolite identification is performed by following a fundamental approach which firstly involves the calculation of elemental composition/molecular formula (MF) of each respective *m/z,* followed by manual searching of an identical MF using online databases, hence identification of the corresponding metabolite. In order to strengthen this identification, authentic standards if available, are then analyzed on the same analytical platform using similar conditions and by comparing either two or more chromatographic and molecular properties such as *m/z*, Rt, maximum absorbance range and fragmentation patterns. This renders or allows conclusive or definitive identification (Brown *et al*., 2011). However, due to unavailability of authentic standards for most natural compounds, an alternative identification strategy is performed. This putative/tentative approach is solely based on chromatographic properties without comparing to standards, but remains a second choice (type 2) to definitive identification as stipulated by the Metabolomics Standards Initiative (Sumner *et al*., 2007).

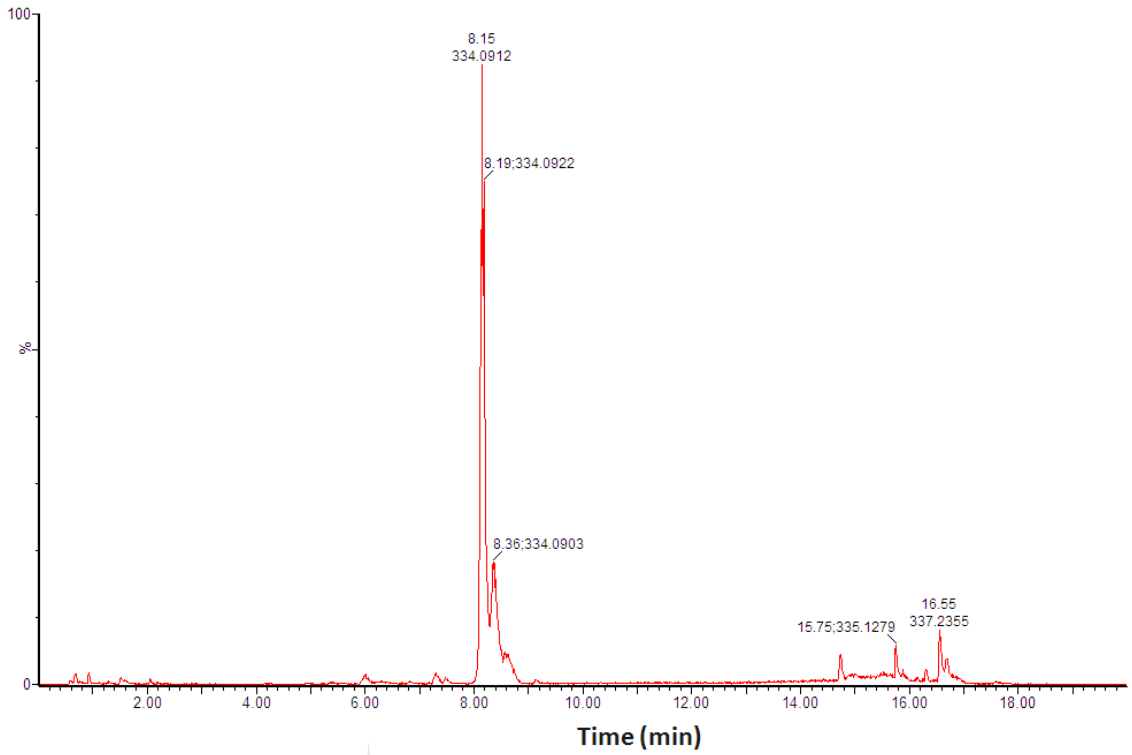


Figure 4: UHPLC-MS/MS chromatogram showing the dominant product (*m/z* = 334.0912) believed to be a sodium-adduct of the bio-transformed product, phenylacetaldoxime-glucoside (structure **6**), derived from INAP. The mass accuracy was 1.5 ppm, and the DBE (double bond equivalent) and iFit (isotopic fit) values were 6.5 and 0.0 respectively.

As previously stated, an untargeted approach was followed in the current study and putative metabolite identification was performed using the PUTMEDID_LC-MS workflow. To the best of our knowledge this is one of the few, if not the first, where this workflow was applied in plant-based metabolomics. As mentioned, metabolite identification by these workflows provides insightful information because the accurate *m/z* and its associated isotopologues are all used to calculate MF. During this process other associated features such adducts, dimers, multiple charged ions and fragments are grouped together with neutral ions during identification. Several examples showing the characteristic results and ability of this workflow to identify metabolites with a high degree of certainty are shown in electronic Supplementary files S1 and S2.

Table 1: List of biomarkers (metabolites) of which the levels were found to be altered (up-regulated) due to INAP treatment.

ID No	*m/z*	Rt	Molecular formula	Adduct	Metabolite name	Ion mode
1	285.0989	7.2	C13H16O7		Benzoyl-glucoside	Pos
1	283.0871	7.2	C13H16O7		Benzoyl-glucoside	Neg
2	121.0309	10.2	C7H5O2		Benzoic acid	Neg
3	313.0931	5.78	C14H17O8		Methoxy-benzoyl glucoside	Neg
4	308.0702	10.86	C10H17NO7	HCOOH	4-Amino-4-deoxyglucuronic acid; α-D-pyranose-form, Me glycoside, *N*-Ac, Me ester	Pos
5	237.0412	10.5	C10H8O4	HCOOH	Scopoletin/5,7-dihydroxy-2-methyl-4H-chromen-4-one	Neg
6	334.0912	8.08	C14H17NO7	Na	Phenylacetaldoxime-glucoside*	Pos
7	340.1028	7.56	C14H17NO6	HCOOH	Prulaurasin ((2R)-(β-D-glucopyranosyloxy)(phenyl)acetonitrile / L-mandelonitrile- β-D-glucoside)	Neg
8	258.1147	11.33	C15H15NO3		2-Hydroxy-1-(2-hydroxyphenyl)-3-phenyl-1-propanone, oxime	Pos

*The elemental composition from the MS/MS also matched that of dhurrin, [(S)-4-hydroxymandelnitrile-β-D-glucopyranoside], (see Fig. 5).

Since the main aim of the current study is to evaluate the effect of INAP on *A. thaliana* plants, much emphasis was placed on metabolites of which the levels were found to be altered due to treatment. Table 1 highlights the metabolites most affected by INAP treatment as predicted by the OPLS-DA based S-loadings plot. From this list, it is notable that most of the metabolites induced are associated with plant defense. For example, benzoic acid (BA) molecules (**1, 2, 3**) are known to play a central role during full implementation of plant defense responses. BA normally exists in its conjugated form (Chong *et al*., 2001). Using HPLC equipped with UV detection, a time dependant accumulation of free and conjugated BA in tobacco undergoing HR after treatment with an elicitor of defense responses were monitored, and this suggest a possible

role during plant defense (Chong *et al.,* 2001). Generally, plant BAs and the respective derivatives are strong mediators of plant responses towards biotic and abiotic stress (Wildermuth, 2006). One notable involvement of BAs is through its unique conjugated form, SA, which has previously been implicated to form the basis of plant defense against wide spectrum of pathogens, and where its accumulation is required for the establishment of local and systemic acquired resistance responses (Dempsey *et al*., 1999). Free and conjugated BAs play significant roles during plant stress responses in general and BA derivatives are also known to have anti-oxidant activities (Natella *et al*., 1999). The accumulation in INAP-treated plants indicates a possible role during the reconditioning of the physiological status of the cells following such an induced stress environment.

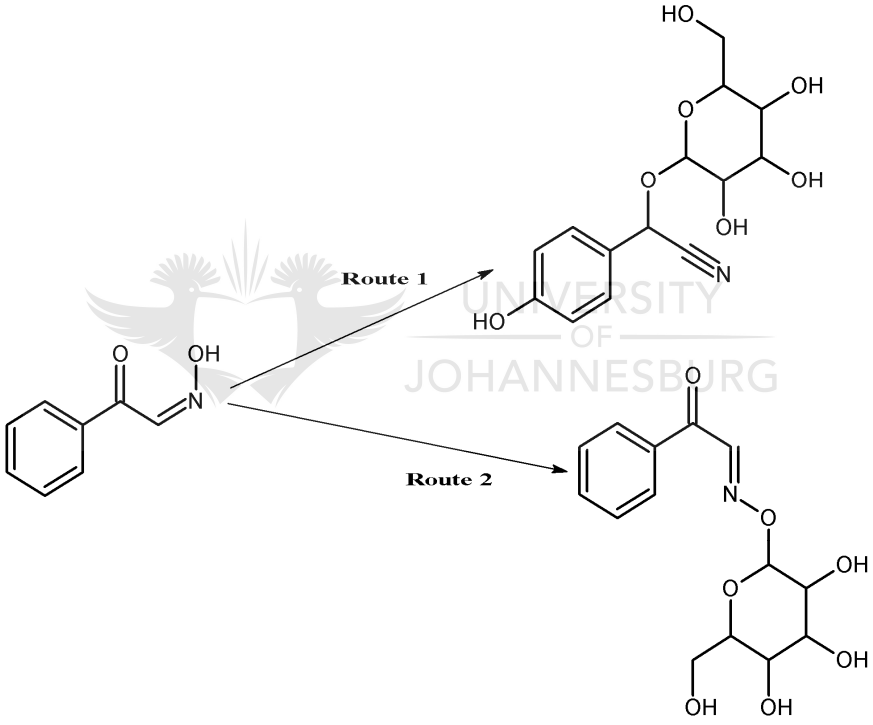


Figure 5: Schematic representation showing two possible hypothetical routes by which INAP can be bio-transformed in *Arabidopsis* plants. Route 1 leads to dhurrin and route 2 to phenylacetaldoxime-glucoside (see also supplementary Figure S1).

Although all of the molecules listed in Table 1 are important, the most interesting identification is (**6**) with the elemental composition ($C_{14}H_{17}NO_7$). According to several online libraries, including the dictionary of natural products (DNP), this molecule was consistently identified as dhurrin, (S)-4-hydroxymandelnitrile-β-D-glucopyranoside. To the best of our

knowledge, dhurrin is only found in sorghum plants (Bak *et al*., 2000) and its accumulation in *A. thaliana* has never been reported. It was from this observation that further experiments were conducted to confirm the identity of this molecule with a mass ion of 334.0912. Initially, MS/MS experiments were designed to monitor only this molecule. However, it was noted that the compound is very susceptible to changes in ionization and the mass ions at *m/z* 312 are only observed under very mild conditions on positive ionization mode. When optimized for the ionization of the mass ion cluster around *m/z* 334, the mass cluster at *m/z* 312 disappears and thus the nominal mass of *m/z* 334.0912 was used. This resulted in only one product (Fig. 4) and the elemental calculation strongly supports the presence of a Na adduct where the computation results yielded $C_{14}H_{17}NO_7Na$, (mass accuracy of 1.5 ppm, isotopic Fit (iFit) value of 0.0 and a double bond equivalent (DBE) value of 6.5). However, when searching for the identity of this molecule, dhurrin was found to be a dominant hit.

Transgenic tobacco and *Arabidopsis* plants expressing two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, became cyanogenic and accumulated metabolites derived from intermediates (*p*-hydroxyphenylacetaldoxime and *p*-hydroxymandelonitrile) in dhurrin biosynthesis (see Supplementary Fig. S1, Bak *et al*., 2000). If the same biotransformation events are considered in this regard, then it is possible that INAP could be converted to dhurrin or phenylacetaldoxime glucoside, and thus the two options as shown in Fig. 5. From the structural properties of the two molecules, it is not convincing that dhurrin can easily form *via* **route 1**, since INAP would first have be hydroxylated and further glycosylated on the carbonyl group, where the latter is not very common for natural products. For the **route 2** option, INAP has to undergo glycosylation on the hydroxyl group of the oxime motif, and this is well supported by the findings of Bak *et al.* (2000).

Oximes are used for the synthesis of glucosinolates and cyanogenic glycosides and are found in plants which are capable of producing such classes of compounds. Citrus plants, however, do not produce these compounds and citaldoxime was thought to serve as a means of protection against oxidative stress induced by ionizing radiation (Dubery *et al*., 1999). In addition, Dubery and colleagues showed that citaldoxime also exhibited anti-fungal activity which was limited to the oxime functional group. INAP, which is a structural moiety of citaldoxime harboring the oxime functional group, was also shown to exhibit anti-fungal activity (results not shown) and can thus be described as an activity-determining factor of citaldoxime.

However, the question regarding the function of oxime metabolites in plants not able to synthesize glucosinolates or cyanogenic glycosides, remains unexplained and it was only recently addressed.

Moller (2010) compared cyanogenic and non-cyanogenic plants based on these systems' respective abilities to metabolize oxime molecules. He further showed that there exists parallel metabolic machinery in such plants, able to metabolize oximes to different products. It was shown that in cyanogenic plants, oximes are metabolized to form glucosinolates and cyanogenic glycoside compounds which function against pest/herbivore attack. However, in non-cyanogenic plants, oximes are metabolized using either one of two routes. In the first case, accumulation of oximes results in mitochondrial dysfunction which subsequently leads to accumulation of reactive oxygen species (ROS) that may act as signal molecules in response to pathogen attack. In the second option, oximes are believed to be detoxified and further metabolized to highly reactive nitroso-compounds which may also give rise to the formation of diverse protein conjugation products that could signal an innate immune response (Moller, 2010). It is conceivable that there is more to the existence of oximes in plants than what was previously thought.

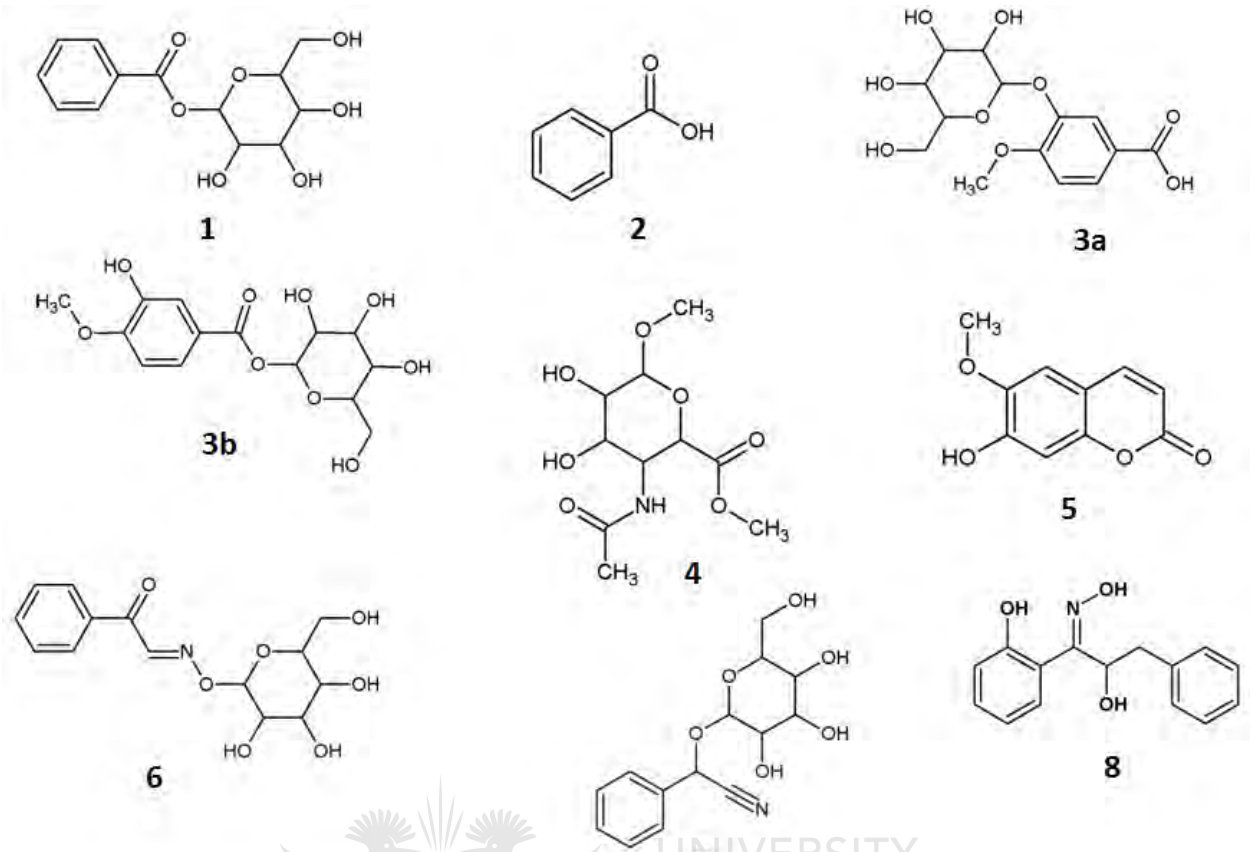


Figure 6: Identified metabolites with the aid of PCA and OPLS-DA loading S-plots from the UHPLC-MS data analyses of INAP treated *A. thaliana* leaf extracts. From the DNP databases markers were identified as :**1**. benzoyl-glucoside, **2**. benzoic acid, **3a**. methoxy-benzoyl glucoside, **3b**. 3-hydroxy, 4-methoxy benzoyl glucoside **4**. 4-amino-4-deoxyglucuronic acid; α-D-pyranose-form, Methyl glycoside, N-Acetyl, Methyl ester, **5**. scopoletin, **6**. phenylacetaldoxime-glucoside, **7**. prulaurasin / L-mandelonitrile- β-D-glucoside, **8**. 2-hydroxy-1-(2-hydroxyphenyl)-3-phenyl-1-propanone oxime .

Although the view point of Moller (2010) can be used to explain the accumulation of citaldoxime in citrus, it was the objective of the current study to further investigate the existence of oximes in non-cyanogenic plants. From the study of Bak and colleagues, where a full metabolic biosynthetic pathway for dhurrin was transferred to *Arabidopsis* and tobacco through genetic engineering, it was shown that the two transgenic plants accumulated cyanogenic glycoside and other related molecules. Amongst these metabolites, *A. thaliana* also accumulated some unusual glucosinolates and it was thus concluded that defense components of one plant can

be transferred to another to result in new metabolite capabilities. This is also the notion from which the current study stemmed.

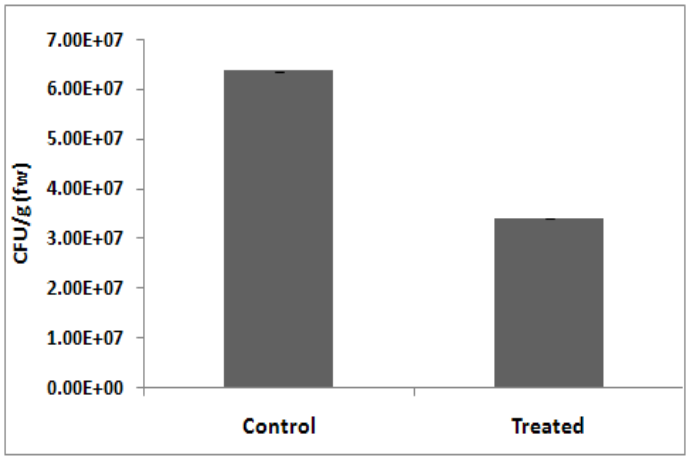


Figure 6: Graphical representation of growth of *P. syringae* pv. *maculicola* in *A. thaliana* leaves after foliar spraying with 1 mM INAP on two consecutive days.

INAP was also found to inhibit bacterial growth *in planta*. When applied to *A. thaliana* plants for two consecutive days, it was found to reduce the growth of *Pseudomonas syringae* pv. *maculicola* (Fig. 7). The protection afforded by INAP treatment serves as an indication that it is capable of inducing an IR response which subsequently induces the priming state. Combining these observations with the metabolite-inducing ability of INAP, it is evident that when applied to plants, this oxime molecule induces both an anti-oxidant and anti-microbial environment *in vivo.*

In conclusion, INAP was found to induce metabolites with known activities towards plant defence responses. The use of UHPLC–MS in conjunction with multivariate data analysis models proved to be successful in the analyses of IR-related metabolites in *A. thaliana* plants. Adaptation of the PUTMEDID_LC–MS workflow for the analysis of plant metabolites was found to ease the identification of several metabolites. Together with our previous and yet to be published findings, the current study presents the unprecedented role of INAP or oximes in plants. Although it is generally assumed that oximes function solely in the biosynthesis of cyanogenic glycosides and glucosinolates, looking at the induced metabolites it can, however, be concluded that oximes are capable of inducing an anti-oxidant environment in plants. This

observation could serve as a justification for the initial accumulation of an INAP–related (citaldoxime) molecule in citrus peel undergoing oxidative stress. INAP was furthermore shown to halt/decrease (approximately 50%) bacterial growth in treated plants, suggesting its possible role in IR-related responses in plants. Adding to our previous finding in which INAP was also shown to undergo biotransformation events similarly to those previously discussed by Bak *et al*. (2000), we conclude that it is possible to use the existing enzymatic machinery of certain plants to produce desired products by means of providing a suitable substrate such as oximes.

6.5. References

Almeida C; Part N; Bouhired S; Kehraus S; König JM (2011) Stachylines A-D from the sponge-derived fungus *Stachylidium* sp. Journal of Natural Products **74**: 21–25.

Bak S; Olsen CE; Halkier BA; Møller BL (2000) Transgenic tobacco and *Arabidopsis* plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. Plant Physiology **123**: 1437–1448.

Beckers GJM; Jaskiewicz M; Liu Y; Underwood WR; He YH; Zhang S; Conrath U (2009) Mitogen-activated protein kinase 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana.* The Plant Cell **21**: 944–953.

Beets CA; Huang J-C; Madala NE; Dubery IA (2012) Activation of camalexin biosynthesis in *Arabidopsis thaliana* in response to perception of bacterial lipopolysaccharides: a gene-to-metabolite study. Planta **136**: 261–272.

Brown M; Wedge DC; Goodacre R; Kell DB; Baker PN; Kenny LC; Mamas MA; Neyses L; Dunn WB (2011) Automated workflows for accurate mass-based putative metabolite identification in LC / MS-derived metabolomic datasets. Bioinformatics **27**: 1108–1112.

Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole-time-of-flight mass spectrometry. Journal of Mass Spectrometry **36**: 849–865.

Chong J; Atanassova R; Werck-Reichhart D; Fritig B; Saindrenan P (2001) Free and conjugated benzoic acid in tobacco plants and cell cultures: induced accumulation upon elicitation of defense responses and role as salicylic acid precursors. Plant Physiology **125**: 318–328.

Conrath U; Beckers GJM; Flors V; Garcia-Augustin P; Jakab G; Mauch F; Newman MA; Pieterse CM; Poinssot B; Pozo MJ; Pugin A; Schaffrath U; Ton J; Wendehenne D; Zimmerli L; Mauch-Mani B (2006) Priming getting ready for battle. Molecular Plant Microbe Interaction **19**: 1062–1071.

Conrath U; Pieterse CMJ; Mauch-Mani B (2002) Priming in plant pathogen interactions. Trends in Plant Science **7**: 210–216.

Coventry HS; Dubery IA (2001) Lipopolysaccharides from *Burkholderia cepacia* contribute to an enhanced defensive capacity and the induction of pathogenesis-related proteins in *Nicotiana tabacum.* Physiological and Molecular Plant Pathology **58**: 149–158.

Dao TTH; Chacon R; Kyong H; Erkelens C; Lefeber AWM; Linthorst HJM; Hae Y; Verpoorte R (2009) Plant physiology and biochemistry effect of benzothiadiazole on the metabolome of *Arabidopsis thaliana*. Plant Physiology and Biochemistry **47**: 146–152.

Dempsey DMA; Shah J; Klessig DF (1999) Salicylic acid and disease resistance in plants. Critical Reviews in Plant Science **18**: 547–575.

Dubery IA; Louw AE; Van Heerden FR (1999) Synthesis and evaluation of 4-(3-methyl-2 butenoxy) isonitrosoacetophenone, a radiation-induced stress metabolite in Citrus. Phytochemistry **50**: 983–989.

Dubery IA, Teodorczuk LG, Louw AE (2000) Early responses in methyl jasmonate-preconditioned cells towards pathogen-derived elicitors. Molecular Cell Biology Research Communications **3**: 105–110.

Dunn WB; Erban A; Weber RJM; Creek DJ; Brown M; Breitling R, Hankemeier T, Goodacre R, Neumann S; Kopka J; Viant MR (2012a) Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics **9**: 44-66.

Dunn WB; Summers A; Brown M; Goodacre R; Lambie M; JohnsonT; Wilkie M; Davies, Topley SN; Brenchley P (2012b) Proof-of-principle study to detect metabolic changes in peritoneal dialysis effluent in patients who develop encapsulating peritoneal sclerosis. Nephrology Dialysis Transplantation **27:** 2502–2510.

Fernie AR; Trethewey RN; Krotzky AJ; Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nature Reviews Molecular Cell Biology **5**: 763–769.

Fiehn O (2002) Metabolomics–the link between genotypes and phenotypes. Plant Molecular Biology **48**: 155–171.

Fiehn O; Kopka J; Dormann P; Altmann T; Trethewey RN; Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nature Biotechnology **18:**1157–1161.

Friedrich L; Lawton K; Ruess W; Masner P; Specker N; Rella MG; Meier B; Dincher S; Staub T; Uknes S; Me´traux JP; Kessmann H; Ryals J (1996) A benzothiadiazole derivative induces systemic acquired resistance in tobacco. Plant Journal **10**: 61–70.

Gatz C; Lenk I (1998) Promoters that respond to chemical inducers. Trends in Plant Sciences **3**: 352–358.

Gerber IB; Laukens K; De Vijlder T;Witters E; Dubery IA (2008) Proteomic identification of cellular targets of lipopolysaccharide-induced signaling in *Nicotiana tabacum* BY-2 cells. Biochimica et Biophysica Acta - Proteins and Proteomics **1784**: 1750–1762.

Gerber IB; Laukens K; Witters E; Dubery IA (2006) Analysis and identification of lipopolysaccharide-responsive phosphoproteins in *Nicotiana tabacum* cells. Plant Physiology and Biochemistry **44**: 369–379.

Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology **43**: 205–227.

Goellner K; Conrath U (2008) Priming: It's all the world to induced disease resistance. European Journal of Plant Pathology **121**: 233–242.

Gomez-Ariza J; Campo S; Rufat M; Estopa M; Messeguer J; Segundo BS; Coca M (2007) Sucrose-mediated priming of plant defense responses and broad spectrum disease resistance by overexpression of the maize pathogenesis related proteins in rice plants. Molecular Plant Microbe Interactions **20**: 363–369.

Gorlach J; Volrath S; Knauf-Beiter G; Hengy G; Beckhove U; Kogel KH; Oostendorp M; Staub T; Ward E; Kessmann H; Ryals J (1996) Benzothiadiazole, a novel class of inducers of systemic acquired resistance, activates gene expression and disease resistance in wheat. Plant Cell **8**: 629–643.

Griffin JL; Nicholls AW; Keun HC; Mortishire-Smith RJ; Nicholson JK; Kuehn T (2002) Metabolic profiling of rodent biological fluids via H^1-NMR spectroscopy using a 1 mm microlitre probe. Analyst **127**: 582–584.

Jolliffe IT (2002) Principal Component Analysis. 2nd edition, Springer, New York.

Jung HW; Tschaplinski TJ; Wang L; Glazebrook J; Greenberg1 JT (2009) Priming in systemic plant immunity. Science **324**: 89–91.

Kim EJ; Kwon J; Park SH; Park C; Seo YB; Shin XHK; Kim HK; Lee KS; Choi SY; Ryu DH; Hwang GS (2011) Metabolite profiling of *Angelica gigas* from different geographical origins using 1 H NMR and UHPLC-MS analyses. Journal of Agricultural and Food Chemistry **59**: 8806–8815.

Kutchan TM (2001) Ecological arsenal and developmental dispatcher: the paradigm of secondary metabolism. Plant Physiology **125**: 58–60.

Lawton KA; Friedrich L; Hunt M; Weymann K; DelaneyT; Kessmann H; Staub T; Ryals T (1996) Benzothiadiazole induces disease resistance in Arabidopsis by activation of the systemic acquired resistance signal transduction pathway. Plant Journal **10**: 71–82.

Lu X; Zhao X; Bai C; Zhao C; Lu G; Xu G (2008) LC–MS-based metabonomics analysis. Journal of Chromatography B **866**: 64–76.

Madala NE; Leone MR; Molinaro A; Dubery IA (2011) Deciphering the structural and biological properties of the lipid A moiety of lipopolysaccharides from *Burkholderia cepacia* strain ASP B 2D, in *Arabidopsis thaliana.* Glycobiology **21**: 184–194.

Madala NE; Molinaro A; Dubery IA (2012) Distinct carbohydrate and lipid-based molecular patterns within lipopolysaccharides from *Burkholderia cepacia* contribute to defense-associated differential gene expression in *Arabidopsis thaliana*. Innate Immunity **18**: 140–154.

Mahadevan S (1973) Role of oximes in nitrogen metabolism in plants. Annual Review in Plant Physiology **24**: 69–88.

Maldonado AM; Doerner P; Dixon RA; Lamb CJ; and Cameron RK (2002) A putative lipid transfer protein involved in systemic resistance signaling in Arabidopsis. Nature **419**: 399–403.

Mishina TE; Zeier J (2007) Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis. Plant Journal **50**: 500–513.

Moller BL (2010) Dynamic metabolons. Science **330**: 1328-1329.

Natella F; Nardini M; Felice MD; Scaccini C (1999) Benzoic and cinnamic acid derivatives as antioxidants: Structure - activity relation. Journal of Agriculture and Food Chemistry **47**: 1453–1459.

Nicholson JK; Lindon JC; Holmes E (1998) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica **29**: 1181–1189.

Oliver SG; Winson MK; Kell DB; Baganz F (1998) Systematic functional analysis of the yeast genome. Trends in Biotechnology **16**: 373–378.

Padidam M (2003) Chemically regulated gene expression in plants. Current Opinion in Plant Biology **6**: 169–177.

Piater LA; Nürnberger T; Dubery IA (2004) Identification of a lipopolysaccharide responsive erk-like map kinase in tobacco leaf tissue. **5**: 331–341.

Pieterse CMJ; Leon-Reyes A; Van der Ent S; Van Wees SCM (2009) Networking by small-molecule hormones in plant immunity. National Chemistry and Biology **5**: 308–316.

Plumb RS; Granger JH; Stumpf CL; Johnson KA; Smith BW; Gaulitz S; Wilson ID; Castro-Perez J (2005) Arapid screening approach to metabonomics using UHPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. Analyst **130**: 844–849.

Ramamoorthy V; Viswanathan R; Raguchander T; Prakasam V; Samiyappan (2001) Introduction to systemic resistance by plant growth promoting rhizobacteria in crop plants against pests and diseases. Crop Protection **20**: 1–11.

Ryals JA; Neuenschwander UH; Willits MG; Molina A; Steiner H; Hunt MD (1996) Systemic acquired resistance. The Plant Cell **8**: 1809–1819.

Ryals J; Uknes S; Ward E (1994) Systemic acquired resistance. Plant Physiology **104**: 1109–1112.

Ryan D; Robards K (2006) Metabolomics: the greatest omics of them all? Analytical Chemistry **78**: 7954–7958.

Sanabria NM; Dubery IA (2006) Differential display profiling of the *Nicotiana* response to LPS reveals elements of plant basal resistance. Biochemical and Biophysical Research Communications **344**: 1001–1007.

Seger C; Sturm S (2007) Analytical aspects of plant metabolite profiling platforms: Current standings and future aims. Journal of Proteome Research **6**: 480–497.

Sumner LW; Amberg A; Barrett D; Beale MH; Beger R; Daykin CA; Fan TWN; Fiehn O; Goodacre R; Griffin JL; Hankemeier T; Hardy N; Harnly J; Higashi R; Kopka J; Lane AN; Lindon JC; Marriott P; Nicholls AW; Reily MD; Thaden JJ; Viant MR (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics **3**: 211–221.

Sumner LW; Mendes P; Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry **62**: 817–836.

Thomma B P H J; Eggermont K; Penninckx IAMA; Mauch-Mani B; Vogelsang R; Cammue BPA; Broe-kaert WF (1998) Separate jasmonate-dependent and salicy- late-dependent defense-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. Proceedings of the National Academy of Science of the United State of America **95**: 15107–15111.

Ton J; Jakab G; Toquin V; Flors V; Iavicoli A; Maeder MN; Metraux JP; Mauch-Mani B (2005) Dissecting the β- aminobutyric acid-induced priming phenomenon in *Arabidopsis*. Plant Cell **17**: 987–999.

Tugizimana F; Steenkamp PA; Piater LA; Dubery IA (2012) Ergosterol-induced sesquiterpenoid synthesis in tobacco cells. Molecules **17**: 1698–1715.

Van der Greef J; Smilde AK (2005) Symbiosis of chemometrics and metabolomics: past, present, and future. Journal of Chemometrics **19**: 376–386.

Van Loon LC; Bakker PAHM; Pieterse CMJ (1998) Systemic resistance induced by rhizosphere bacteria. Annual Review in Phytopathology **36**: 453–483.

Vicedo B; Flors V; De la; Oleyva M; Finiti I; Kravchuk Z; Real MD; García-Agustín P; González-Bosch C (2009) Hexanoic acid-induced resistance against *Botrytis cinerea* in tomato plants. Molecular Plant-Microbe Interaction **22**:1455–1465.

Von Roepenack-Lahaye E; Degenkolb T; Zerjeski M; Franz M; Roth U; Wessjohann L; Schmidt J; Scheel D; Clemens S (2004) Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. Plant Physiology **134**: 548–559.

Walters DR; Paterson L; Walsh DJ; Havis ND (2009) Priming for plant defense in barley provides benefits only under high disease pressure. Physiological and Molecular Plant Pathology **73**: 95–100.

Wildermuth MC (**2006)** Variations on a theme: synthesis and modification of plant benzoic acids. Current Opinion in Plant Biology **9**: 288–296.

Wiklund S; Johansson E; Sjo L; Shockcor JP; Gottfries J; Moritz T; Trygg J (2008) Visualization of GC / TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical Chemistry **80**: 115–122.

Zhang SJ; Yang X; Sun MW; Sun F; Deng S; Dong HS (2009) Riboflavin-induced priming for pathogen defense in *Arabidopsis thaliana.* Journal of Integrative Plant Biology **51**: 167–174.

Zimmerli L; Metraux JP; Mauch-Mani B (2001) β-aminobutyric acid-induced protection of Arabidopsis against the necrotrophic fungus *Botrytis cinerea*. Plant Physiology **126**: 517–523.

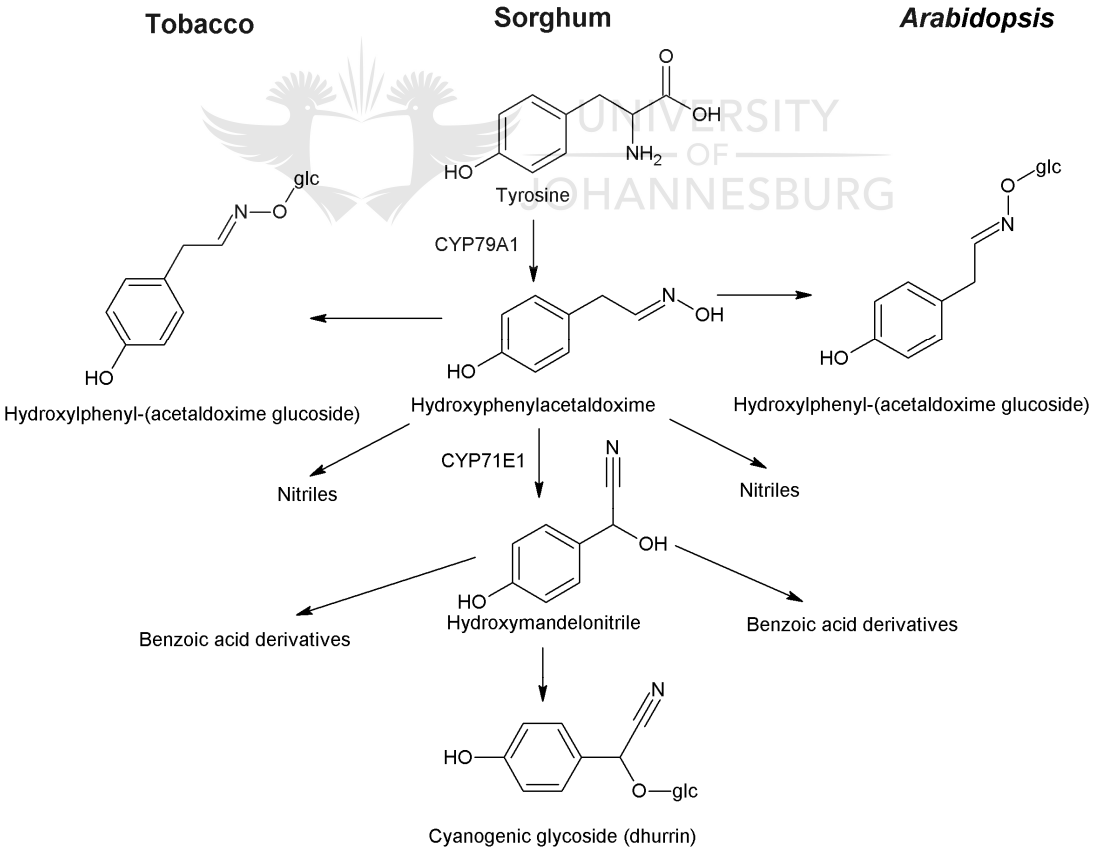
Zimmerli L; Jakab C; Metraux JP; Mauch-Mani B (2000) Potentiation of pathogen-specific defense mechanisms in *Arabidopsis* by β-aminobutyric acid. Proceedings of the National Academy of Science of the United State of America **97**: 12920–129

6.6. Supplementary files

Electronic Supplementary File S1: List of metabolites identified with the aid PUTMEDID_LC-MS workflow using negative ionization UHPLC-MS data, see CD attached.

Electronic Supplementary File S2: List of metabolites identified with the aid PUTMEDID_LC-MS workflow using positive ionization UHPLC-MS data, see CD attached.

Supplementary Fig. S1: Schematic representation showing the route by which sorghum plants produce the cyanogenic glycoside, dhurrin. The branching points indicate possible routes which other non-cyanogenic plants such as *Arabidopsis* and tobacco can follow in the presence of the precursor molecule *p*-hydroxyphenylacetaldoxime (Bak *et al*., 2000). By providing a related oxime molecule (INAP), to both *Arabidopsis* and tobacco, similar results were obtained during the current study.



Chapter 7: Deciphering time-dependent trends in metabolomic data from elicited plant cells using multivariate statistical models

7.1. Abstract

Metabolomics is a systematic study of the unique chemical fingerprints that specific cellular processes generate. Metabolomic analyses result in high dimensional data files which are cumbersome to analyze manually, requiring tools capable of handling complex data matrices. Chemometrics-based models that have been successfully applied for the analyses of metabolomics data include principal component analysis (PCA) and linear regression models such as orthogonal projections to latent structures discriminant analysis (OPLS-DA). As an unsupervised model, PCA allows patterns within the data to be revealed by reducing the dimensionality thereof into small variables which are easier to explain. Here, PCA-based hierarchical cluster analysis (HCA) and metabolic trees as alternative methods for data visualisation are exploited in parallel to reveal the time-dependent changes in the metabolite profiles of sorghum and tobacco cells treated with the xenobiotic, isonitrosoacetophenone (INAP). HCA and metabolic trees were able to show that INAP induces metabolic perturbations in both cell lines and that cellular homeostasis is re-established over time. OPLS-DA-based shared and unique structure (SUS) plots were also utilized to confirm the results obtained from both HCA and metabolic trees. Although similar, the response in sorghum cells was found to be more consistent and well coordinated compared to tobacco cells.

Keywords: Nicotiana tabacum; Sorghum bicolor; Metabolomics; Isonitrosoacetophenone; PCA; HCA; OPLS-DA; SUS; Metabolic trees.

7.2. Introduction

Metabolomics is an unbiased technique aimed at measuring the metabolite content of a cell, tissue or organism under a given physiological status [1, 2]. It is the analyses of these metabolites which lead to a comprehensive understanding of the unique chemical fingerprints that result from specific cellular processes [3] and, as opposed to the analysis of genes or proteins, allows a thorough elucidation of the phenotypical characteristics of living systems. Metabolomics has recently found significant applications in many fields such as responses to environmental stresses [4,5], studying global effects of genetic manipulation, nutrition and health [6-9] and, most importantly, in plant studies [10-13].

A few steps are mandatory for metabolomic studies and these include sample preparation, choice of analytical platform, and subsequent data analyses. In sample preparation, the choice of solvents and method is defined by the physico-chemical properties of the metabolites and the anticipated outcomes of the study. The selection of solvents and suitable pH should be well optimized before any experimental procedure is carried out [13]. The choice of analytical technology is also as important as sample preparation. Here, techniques such as gas- and liquid chromatography coupled to mass spectrometry (GC-MS and LC-MS) [14, 15] and nuclear magnetic resonance (NMR) spectroscopy [13] are commonly used. More comprehensive details on these and other techniques have previously been reported in metabolomics [13, 16, 17]. In the current study, UHPLC-MS was used for metabolite data acquisition based on its technological advances and ability to analyze a broad spectrum of metabolites of different polarities [18, 19]. Using similar analytical conditions, UHPLC-based methods detect more metabolites than traditional HPLC-based methods and thus generates more data output [20]. Depending on the analytical platform and the method of data analysis, different data outputs can be obtained and if

care is not taken, these may wrongly be interpreted as biological variation whilst being due to experimental variation.

Data analysis is an essential step during metabolomic studies, since meaningful information needs to be extracted from structurally complex datasets [21]. It is therefore important that the design of metabolomic experiments is well standardized so that valid and reproducible results can be converted into biological knowledge.

Biotransformation is the process in which one chemical is transformed into another by means of a biocatalyst or biological system, and includes metabolism of xenobiotic compounds [22]. Novel enzyme-substrate combinations *in vivo* can lead to the biosynthesis of new, natural product-derived compounds [23]. We have previously reported that isonitrosoacetophenone (INAP), a structural analogue of citaldoxime, a phytoalexin and anti-oxidant stress metabolite [24,25], is metabolized and bio-converted in tobacco cells [26]. In the current study, comprehensive data analyses with the aid of multivariate data analysis (MDA) models such as Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and the Shared and Unique Structures (SUS) plot generated by Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), were used to investigate the effect of INAP in tobacco as well as sorghum cell suspensions. HCA, SUS plot and metabolic trees [27] were used together for the first time to decipher the time-dependent responses in the two cell lines which allowed comprehensive differentiations to be drawn with regard to the metabolism of oximes.

7.3. Materials and methods

7.3.1. Cell culture, treatment and metabolite extraction

Nicotiana tabacum cv. Samsun and *Sorghum bicolor* L. Moench (sweet white) cell suspensions were cultured as previously described [44, 45]. Three days after sub-culturing, aliquots (20 mL suspensions) were treated with 250 mM isonitrosoacetophenone (INAP), dissolved in acetone, to a final concentration of 1 mM with continuous rotation at 80 rpm and 25ºC for 6, 12, 18, and 24 h time intervals. The final concentration of acetone was 0.4% and control cells received no treatment. For the experimental design, a minimum of ten replicates for each biological group was utilized. After elicitation, cells were collected by means of vacuum filtration and metabolites extracted from the wet cells (2 g) by homogenization in 1 mL 100% methanol. To aid maximum extraction, the homogenates were allowed to agitate on a rotary shaker for at least 1 h. Cell debris was removed by means of centrifugation at 5000 *x* g for 10 min. The resulting supernatant was transferred to a new tube and the volume reduced to 1 mL with the aid of a Buchi rotary evaporator operating at 45ºC, followed by drying to completeness in a 2 mL microcentrifuge tube using a speed vacuum centrifuge operating at 45ºC. The resulting pellet was dissolved in 400 µL 50% methanol and filtered through a 0.22 µm filter into a new UHPLC glass vial fitted with a 0.1 mL insert.

7.3.2. Chromatographic- and Mass spectrometric conditions

Chromatographic conditions were adapted from our previous work [26]. Briefly, methanol extracts (5 µl) were analyzed on a UHPLC connected synapt high definition MS instrument (Waters, Corporation, USA) equipped with an Acquity BEH C18 column (100 mm × 2.1 mm with particle size of 1.7 µm) (Waters Corporation, USA). Two technical replicates for 5

independent samples were performed resulting in 10 injections for each biological group (control, 6, 12, 18, and 24 h). The composition of mobile phase A consisted of 0.1% formic acid in deionized water and mobile phase B consisted of 0.1% formic acid in methanol. The column was eluted with a linear gradient at a constant flow rate of 400 µl min^{-1} of 5% B over 0.0-2.0 min, 5-95% B over 2.0-22.0 min, held constant at 95% B over 22.0-25.0 min, 95-5% B over 25.0-27.0 min and a final wash at 5% B over 27-30 min. For MS acquisition, data was collected on a centroid mode and negative polarity electro-spray ionization (ESI) with a collision energy of 3 eV. Instrumental settings were as follows; capillary voltage: 2.5 kV, sample cone voltage: 17 V, extraction cone voltage: 5.0 V, MCP detector voltage: 1600 V, source temperature: 120°C, desolvation temperature: 350°C, cone gas flow: 50 (L h^{-1}), desolvation gas flow: 450 (L h^{-1}), *m/z* range: 100-1000, scan time: 0.1 sec, interscan delay: 0.02 sec, lockmass: leucine enkephalin (556.3 g/mol), lockmass flow rate: 0.4 mL min^{-1}, mass accuracy window: 0.5 Da.

7.3.3. Data analyses

Primary data was further analyzed by MarkerlynxXS™ software (Waters Corporation, Milford, USA) with parameters as follows: retention time (Rt) of 1-27 min, mass range of 100-1000 Da, mass tolerance of 0.02 Da, Rt window of 0.2 min and, furthermore, isotopic peaks were excluded from the analysis. Data was normalized to total intensity (area) using Markerlynx. The datasets thus obtained were exported to the SIMCA-P software version 12.0 (Umetrics, Umea, Sweden) in order to perform PCA and OPLS-DA. Before performing these multivariate data analyses (MDA), data was mean centered and Pareto-scaled for both models.

In order to evaluate the effect of time on the response, HCA was automatically calculated and the resulting dendrogram evaluated with the aid of the SIMCA-P software. For HCA analysis, the Ward distance algorithm was used to calculate the distance between the different

generated clusters. Using the PCAtoTree programme [27], the metabolomic tree diagrams were created and the corresponding bootstrap values calculated to interpret the PCA clustering pattern. Unlike in the case of HCA, where the Ward method was used, these tree diagrams were generated using the Euclidean distances method between the clusters from the PCA scores plots (Fig. 1). Here, the standard bootstrapping techniques were used to generate a set of 100 distance matrices by randomly re-sampling the cluster centers and Euclidean distances. The matrices were then used in the PHYLIP phylogenetic software package (http://www.phylip.com) [46] to generate 100 tree diagrams and a consensus tree diagram. The numbers on the trees indicates the bootstrap values which describes the number of times each node was present in the set of 100 tree diagrams. Bootstrap numbers below 50% indicates insignificant separation between the clusters.

7.4. Results and discussion

In the current study, two metabolically distinct cell lines from tobacco (*Nicotiana tabacum*) and sorghum (*Sorghum bicolor*) plants were treated, for different time intervals, with an oxime molecule (INAP) which represents a precursor/activity determining motif of citaldoxime, a molecule with known anti-oxidant and antifungal activities [25]. Metabolites extracted from these cells in response to the treatment were analyzed using UHPLC-MS and subsequent data analyses were performed. As previously mentioned, metabolomic studies result in highly complex data which are spread in high-dimensional planes, and the dimensionality reduction is an important first step for pre-processing such data [28]. Traditionally, reduction techniques such as PCA, and multivariate data models like PLS-DA and OPLS-DA, are used to achieve this [28]. In the current study, PCA, HCA, metabolic trees and OPLS-DA models were exploited for data

visualization and useful information, such as the underlying biological responses of tobacco and sorghum cell suspensions towards INAP, was deduced.

7.4.1. Principal Component Analysis

By mathematical definition, PCA is an orthogonal linear transformation of possibly correlated variables into a smaller number of uncorrelated variables called principal components (PCs), where the greatest variance within the data by any projection is explained on the first co-ordinate (PC1) and the least variance is explained/projected by subsequent PCs [29]. PCA and other reduction models thus convert the data obtained by high-throughput instruments into a simple visual representation known as score plots which show the data as the clustering of biological samples into either similar or different groupings. Here, sample data from different biological backgrounds are clearly separated into distinct clusters and samples that cluster together can be referred as a specific "metabolic phenotype" [30].

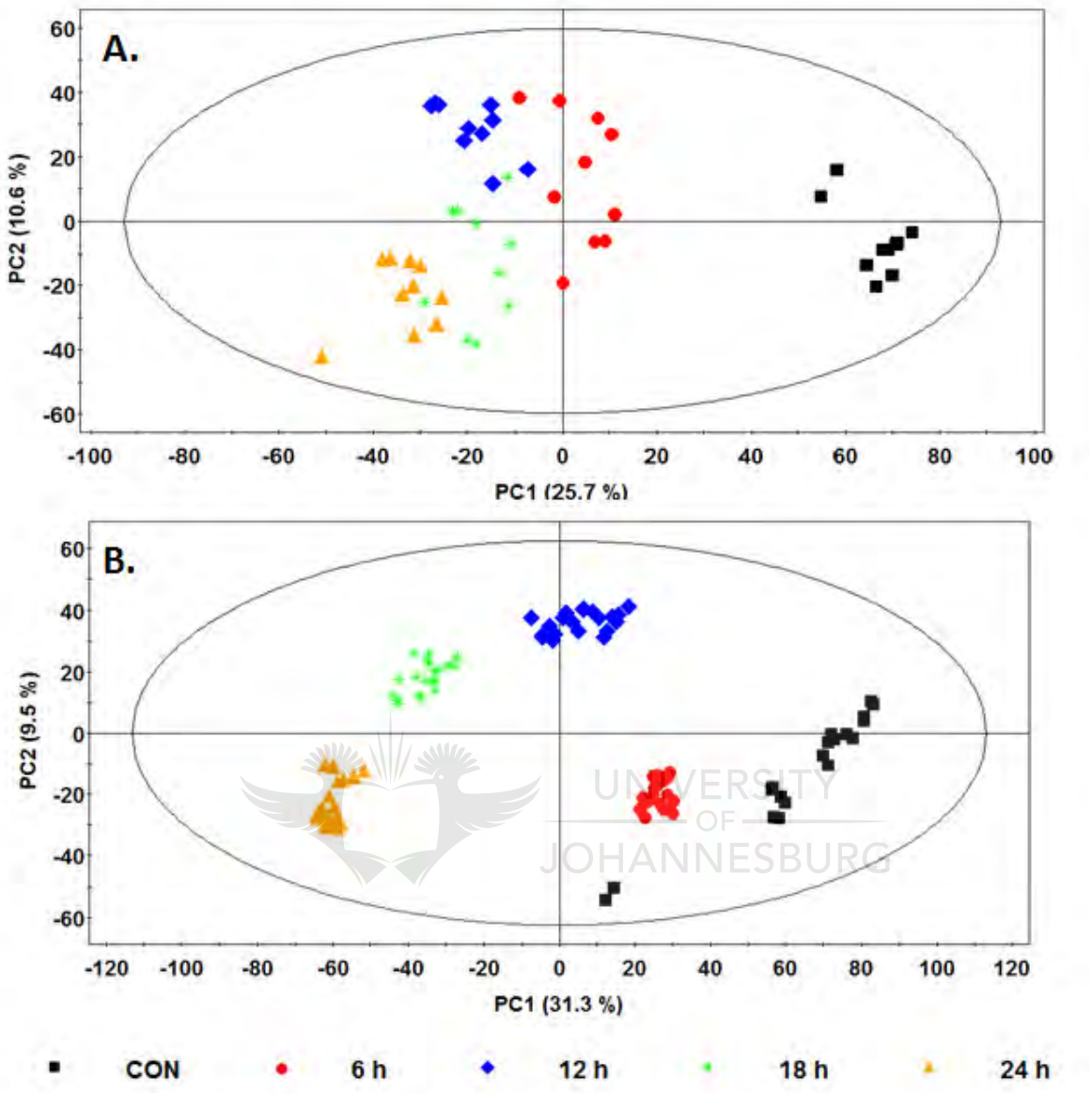


Figure 1. PCA score plots showing the different clusters of samples from tobacco (**A**) and sorghum (**B**). Mid-polar metabolites were extracted from INAP-treated tobacco and sorghum cell suspensions at different time intervals as represented by different colors and symbols on the plot (key for different time intervals is indicated). Model validation gave $R^2X = 0.6$ and $Q^2_{(cum)} = 0.50$ for the tobacco PCA model (4 PCs) and $R^2X = 0.64$ and $Q^2_{(cum)} = 0.48$ for the sorghum PCA model (7 PCs).

From the PCA score plots (Fig. 1), it can be seen that INAP was capable of inducing metabolic perturbations in both cell lines. The samples originating from the treated and non-treated cells clustered in different areas in the plots. As expected, the plot shows that variation between the different biological/treatment groups is more pronounced on the PC1 which constitutes the highest variation in the models. The corresponding PC1 (describing the variation between groups) from the two plots was found to be 25.7% and 31.3% respectively, and PC2 (which describes the variation within the groups) was 10.6% and 9.5% for the tobacco and sorghum models respectively. However, the difference amongst all the treatment time intervals was found to be not as distinct, especially when the later time points (12, 18, and 24 h) are considered. From these plots (Fig. 1A and B) it is clear that although the 6 h time point exists as a distinctive cluster, it possesses less variation from the control as compared to the other time points and could be due to the fact that the metabolic responses are still minimal at such an early time interval. These trends cannot be deduced when later time points are also considered, thus the response appears to progress non-linearly from control to 6 h to 12 h, and so on. In cases such as the one where variation amongst the later time points is less prominent (due to the fact that the separation of data clusters is not as clearly defined), measures need to be taken to overcome this. It should be noted that even though the presentation of data in PC scores space is the result of a statistical analysis, this only shows a qualitative separation and the degree of separation between data clusters is not quantitatively addressed by the score plots [27]. Thus, the basic statistical question regarding significant differences between the clusters is not addressed by PCA score plots even though the visualization represents a qualitative clustering due to metabolic differences. As such, subsequent plots, including the loadings plot, are used to evaluate the causative factors which result in different clustering on score plots [31]. From here,

metabolites that are either up- or down-regulated can be selected to further evaluate the degree of significance across the clusters which they influence. The loading plots (not shown) corresponding to these PCA score plots revealed metabolites which can be assumed to be influential of the clustering seen on the score plots and that were unique in the two plants.

From the results it is evident that PCA score plots suffice the understanding of 'superficial' clustering/separation of samples due to the respective biological/biochemical background. However, PCA is not capable of showing the underlying degree of similarities between the different clusters and hence the trend of responses within the data.

7.4.2. Hierarchical Cluster Analysis

HCA was used as a complimentary model for data visualization. HCA is another data reduction method which is mainly used for finding the underlying structure of objects through a repetitive process that associates (agglomerative methods) or dissociates (divisive methods) object by object until all are equally and completely processed [32]. PCA results in score plots for data visualization, while HCA usually produces a dendrogram, or tree-like diagram, as final output [32-34]. In the current study, a fully automated HCA was performed on the data, similarly to PCA generation as described in 2.1. The resulting dendrogram was calculated using the Ward method (Ward, 1963). Here, the Ward linkage method states that the distance between two clusters is described by how much the sum of the squares will increase when the two are merged (Sato *et al*., 2008). The distance between the clusters is given by the following formula:

$$d(a,b) = \frac{n_a n_b}{n_a + n_b}(x_a - x_b)^2$$

where ***d*** is the distance between cluster ***a*** and ***b***, ***x*** is the center of each cluster and ***n*** is the number of points in each respective cluster [35, 36].

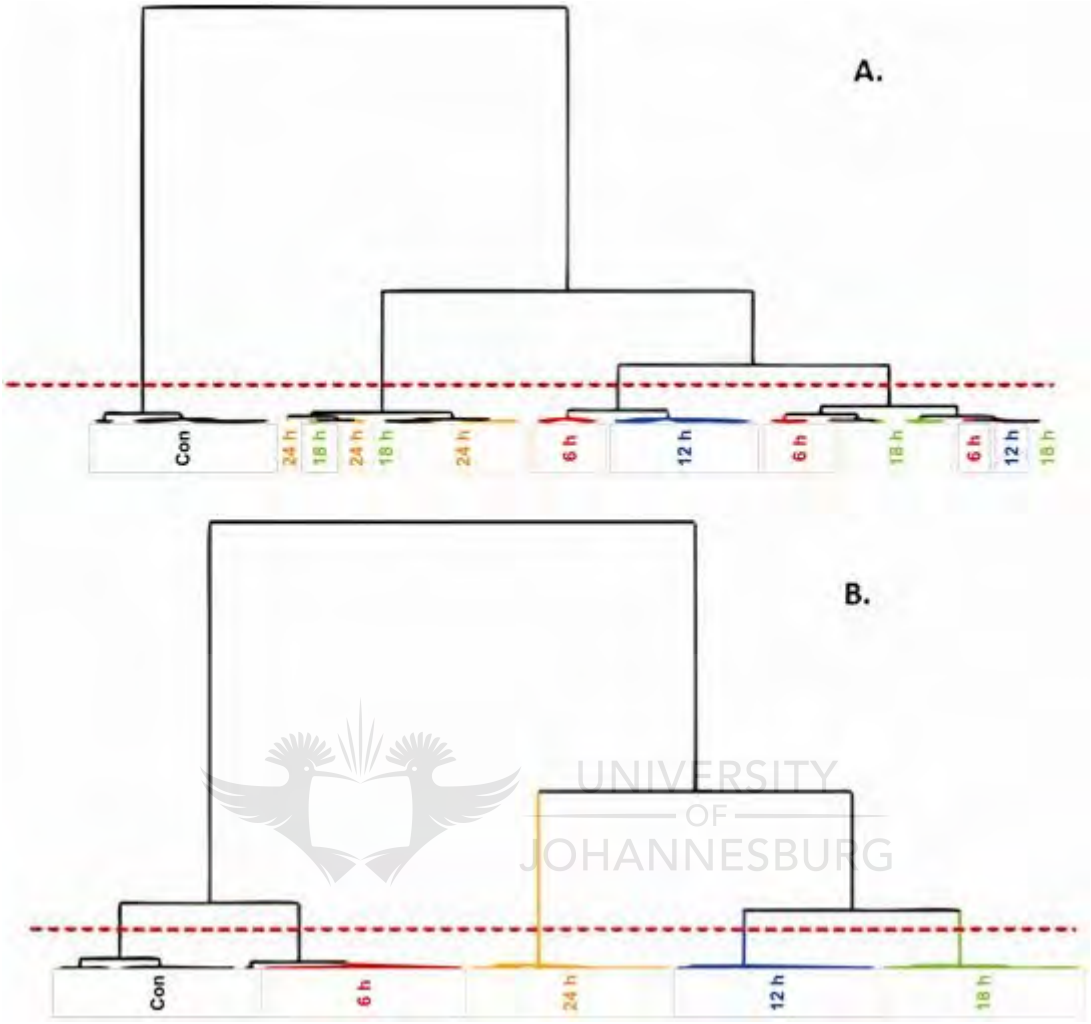


Figure 2. HCA dendrograms showing the relationship between samples originating from INAP treated tobacco (**A**) and sorghum (**B**) cells at different time intervals. The plot shows the relation between samples as described by the length/distance of the node linking two clusters. The number of clusters can be deduced by counting the regions in which the red dotted line crosses the node of each respective cluster. The coloring of individual samples in the dendrogram correlate with colored symbols in the PCA scores plot.

The HCA dendrograms (Fig. 2) show that, although similar results to those of PCA were obtained, more comprehensive findings can be deduced. Taking the tobacco HCA results (Fig. 2A) into account, a definitive clustering among the control samples can be seen. When the different treatment times are however considered, no definitive clustering exists and samples from the same biological group (treatment times) are spread across different clusters. From these results, if a line is placed horizontally across the clusters as shown by the red dotted line on the plot, a maximum of four distinctive clusters can be identified. The first cluster exclusively contains all the control samples. The second cluster is dominated by the 24 h samples and also contains some samples from the 18 h treated time point. The third cluster is dominated by the 12 h treatment as well as some traces of samples from 6 h, and lastly, equal amounts of samples from 18 h and 6 h are seen in the fourth cluster which also contains a few samples from 12 h time interval. These results are clearly different from that seen on the PCA score plots.

Interestingly, the results obtained with sorghum samples show a very well structured response due to INAP treatment unlike tobacco, where maximum variation only exists between the control group and treatment samples as a whole. In sorghum, a well consolidated response exists which can be explained more comprehensively, and underlying this response is a time factor which is more pronounced in the HCA dendrogram. Similarly to the tobacco case, if a line is considered in approximately the same position as in the tobacco HCA, it can be seen that five clusters exist. These depict the fundamental biological/treatment groups (control, 6 h, 12 h, 18 h, and 24 h) which were originally used in the study. From the results, the first cluster exclusively contains samples from the control group, the second cluster contains samples from 6 h, the third contains samples from 24 h, the fourth cluster contains 12 h and the fifth group contains samples

exclusively from the 18 h treated time point. These results are indicative of more stringent metabolism of INAP by the sorghum cells in comparison to the tobacco cells.

Sorghum is a cyanogenic plant which is able to metabolize oxime containing precursors [37]. As previously mentioned, INAP is an oxime molecule similar to intermediates/precursors during the biosynthesis of glucosinolates and cyanogenic glycosides. These two classes of molecules play vital roles during plant: herbivore interactions [38]. Plants capable of metabolizing oxime precursors that are subsequently used for defense responses include sorghum, Arabidopsis [37] and cassava [39], but not tobacco. The existence of oximes in non-oxime metabolizing plants has been reported, e.g. the induced synthesis of citaldoxime in citrus flavedo in response to gamma-radiation treatment [25]. These findings suggest a possible role in plants other than the one stated above. The enzymes which code for the synthesis of cyanogenic glycosides and other related compounds exist in a tightly associated complex known as metabolons. In other plants, the same set of enzymes might exist as well, but are found to be a loosely associated metabolon and sometimes not all are present. In the latter case, oxime precursors do not result in the accumulation of cyanogenic glycosides or glucosinolates [37].

The genes encoding the defense components of one plant can be transferred to another to result in new metabolite capabilities [37]. In contrast to transgenic approaches, new metabolites can also be generated by supplying precursor molecules that are capable of being recognized by enzymes already present in the plant [26]. The findings of the current study, as shown by the HCA, suggests that the metabolic machinery of sorghum cells recognizes the oxime molecule more efficiently than that of tobacco cells, which shows variability across the different treatment time intervals.

7.4.3. Metabolic tree and bootstrapping

An advanced data visualization model, metabolic tree diagrams (PCAtoTree programme), was also used to process the primary data. Unlike the HCA, these trees offer significant advantages as they represent data that are statistically justified [27]. This is due to the fact that during the generation of these trees, bootstrapping numbers which determine the significant differences between the different clusters are also shown [40]. From the current study, two independent tree diagrams corresponding to the tobacco and sorghum samples were generated using the data from the 2D PCA score plots.

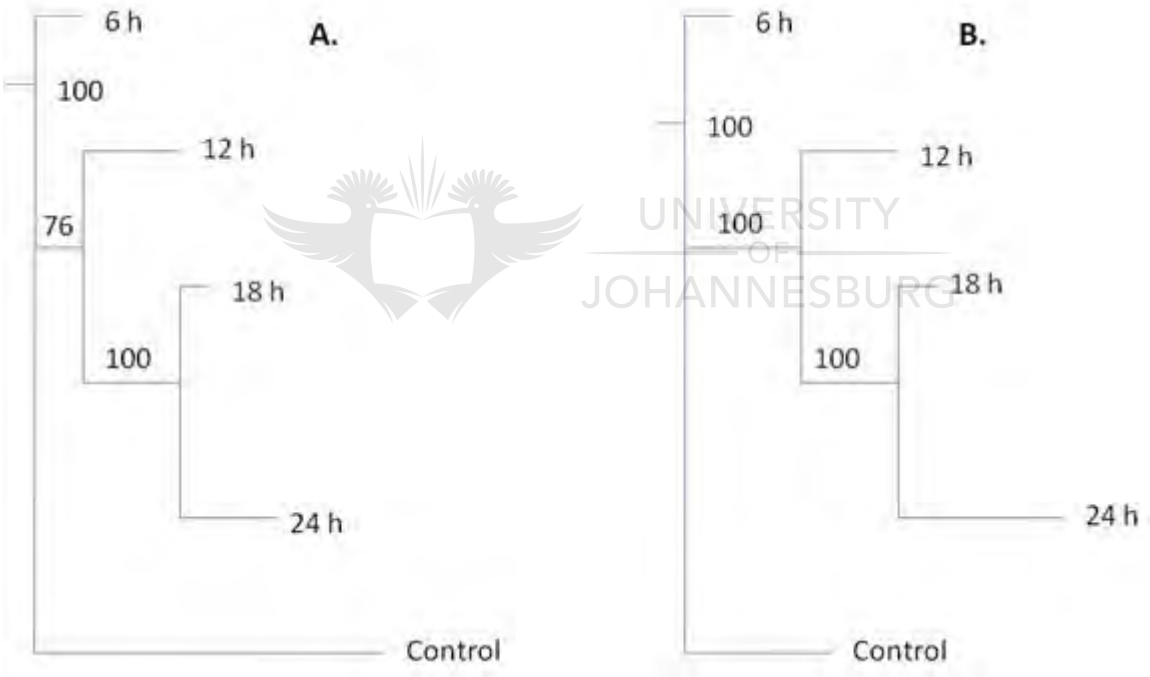


Figure 3. Metabolomics tree diagram determined from the PCA scores plots of both tobacco (**A**) and sorghum (**B**) samples. The bootstrap numbers for each node are indicated on the tree diagram.

When comparing the trees (Fig. 3), the two diagrams appear similar, but with close visual inspection it can be seen that the distance between the clusters are different. This is due to

different metabolic responses exhibited by the two plant systems. Similarly to the HCA, the length of the lines (node) connecting the groups describes the distance between the clusters; for instance the horizontal line connecting the control and the rest of the groups is longer on the tobacco tree than it is on the sorghum tree. This is due to the fact that there is larger distance between the control group and the treated group on the tobacco data than for the sorghum system. Still, on both tree diagrams it is notable that the 18 h and 24 h are closely related to each other, and this cluster appeared 100 times in both cases. The 18 h/24 h cluster is more closely related to the 12 h cluster than it is to the control and 6 h clusters. However, the cluster of 12 h/18 h/24 h appeared 100 times in sorghum and only 76 times on the tobacco system. This is consistent with the PCA results as it can be seen that the distance between the treatment times is smaller in tobacco than it is in the sorghum system. In deciphering the trends of these tree diagrams, it can be observed that the control is more closely related to the 6 h, followed by the 24 h/18 h and then lastly to the 12 h cluster. The fact that the bootstrapping values in the sorghum system were always 100 is evidence that there is a definite separation between the groups that is not as clear in the tobacco system. These observations are in line with those seen on the HCA.

7.4.4. Shared and Unique Structure (SUS) plots

It is clear that PCA only evaluates global patterns (maximum variation) within the data and is not a good tool for revealing local phenomena. For the same and other reasons stated by Van der Greef and Smilde (2005), alternative techniques have been proposed. Here, a supervised model, OPLS-DA, was used to reveal underlying responses which is associated with a time-trend as shown by the HCA above. OPLS-DA can be considered as an extension of the traditional

PLS-DA, and it was proposed solely to handle the class orthogonal variation (hence the name) in the data which is not explained by PLS-DA [31, 42, 43].

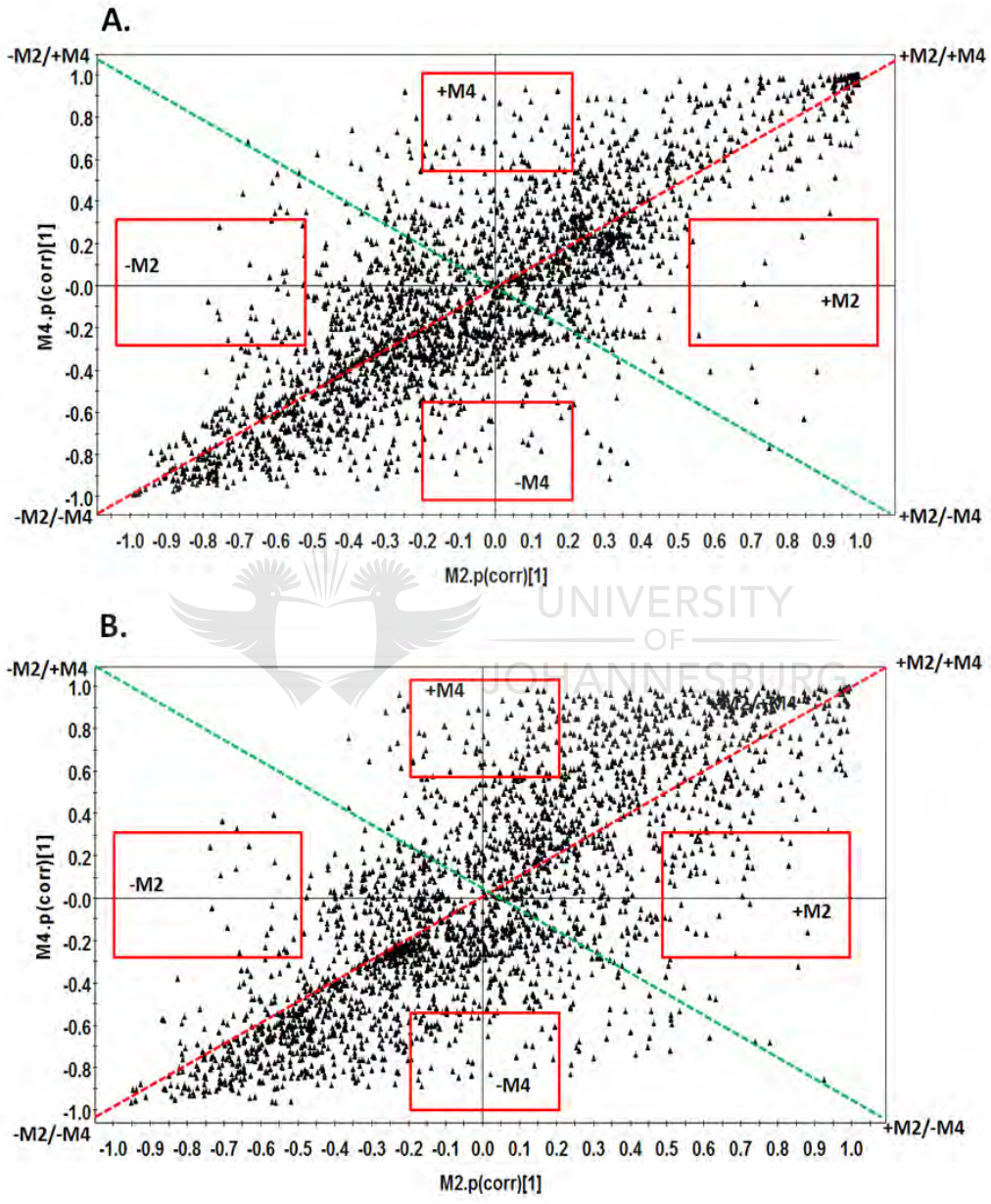


Figure 4. The representative SUS-plots constructed using a two correlation coefficient (p[corr]) from two independent OPLS-DA loadings S-plots (Control vs. 6 h (M2) and Control vs. 18 h (M4)) for both tobacco (**A**) and sorghum (**B**). This plot shows how the metabolites from one

independent model relates to those from the corresponding model. The regions in which shared and unique metabolites reside are highlighted on the plot. The description of the different regions is given in the main text.

In detail, PLS-DA is a linear regression model which seeks to find relationships between two respective data tables (***X*** and ***Y***), where ***X*** is normally the instrument derived/measured data (i.e. GC/LC-MS or NMR data) while ***Y*** represents a binary vector which is associated with class membership. OPLS-DA separates the variation on ***X*** into two parts, one that is linearly related to ***X*** and another one that is orthogonal to ***Y***. This fragmentation on the X variation results in OPLS-DA comprised of two variations, one which is the ***Y***-predictive variation and explains the variation between class membership, and the other which is the ***Y***-orthogonal, which explain the variation within the class membership [31].

In the current study, OPLS-DA was used to highlight the variables responsible for differences among the various groups (classes represented by different time intervals). Similarly to PCA, score plots are also used to evaluate the differences in clustering between samples from different biological groups or classes in OPLS-DA. However, unlike PCA, the loadings plots for OPLS-DA contains statistical sound parameters which explain the underlying statistical significance that forms the basis of separation/clustering of different group on the score plots. OPLS-DA is therefore a very good model to identify variables (metabolites) of which the levels are perturbed due to a certain treatment as compared to control groups. This observation is made possible by plots such as the S-plots which are calculated to visualize the relationship between co-variance and correlation within the OPLS-DA results and to reveal the statistically significant

metabolites [31]. The SUS plot is another visualization plot proposed by Wiklund and associates. The SUS-plot combines the correlations from two different OPLS-DA models, and facilitates the identification of shared and unique metabolites from the two respective treatments with a common control. The SUS plot is capable of showing the metabolites which are unique in one model (as increasing or decreasing) and is also capable of showing those which are shared/correlated between the two models. This correlation can be positive (increasing/decreasing in both models) or negative (decreasing in one and increasing in another model).

In the current study, the use of SUS-plots was adapted to decipher the time-dependent response associated with different time intervals and, according our knowledge, this is the first report of its kind on any biological data. Basically, OPLS-DA models are generated by comparing control and treated samples represented by each time interval. From each cell line, a maximum of six (6) different plots (6 h vs. 12 h, 6 h vs. 18 h, 6 h vs. 24 h, 12 h vs. 18 h, 12 h vs. 24 h, 18 h vs. 24 h) were generated and compared to each other. These combinations were derived from the respective loadings S-plots generated from the four different models, Control vs. 6 h (M2), Control vs. 12 h (M3), Control vs. 18 h (M4), Control vs. 24 h (M5) (data not shown). From the results, the SUS-plot was found to be more complementary to the HCA, since the same pattern can also be drawn from both. Fig. 4 shows the SUS-plots generated by comparing M2 and M4 (6 h vs. 18 h), from both sorghum and tobacco. These two time points were chosen as they represent different stages of responses: the 6 h (M2) represent an early response and 18 h (M4) represent a mid to late response. Here, it can be seen that metabolites in the tobacco case are more positively correlated and less scattered than in the sorghum case at the same time points. In order to draw such a conclusion, one should consider the "key"

symbols/signatures provided on the plots. For instance, all the metabolites scattered across the red dotted line represent those which are positively correlated (++/--). Metabolites which are scattered across the green dotted line represent those which are negatively correlated (+-/-+). In the second case, those which are found on the red boxes across the plot axes are either increasing (+M) or decreasing (-M) for that particular model and, unlike the former case which describes the "shared structures", the latter cases describes the "unique structures". It is also important to note that metabolites which are on the extreme ends (outliers) of the axes contribute more significantly than those close to the center. Still, on the M2 vs. M4 tobacco SUS-plot, it can be seen that the distribution of the metabolites seems to create a latent line across the positively correlated diagonal line. The same is seen in the case of sorghum but is less pronounced as more metabolites are spread all over the plot, especially on the positive side of both M2 and M4. It is such spreading which shows less "sharing" of metabolites between the two models, hence a different/distinctive metabolic phenotype. When all the SUS-plots are considered (Supplementary figures S1 and S2), it can be seen that there is a tighter distribution of metabolites across the different time points on the tobacco cell line than in the sorghum system. This observation strongly supports the hypothesis that oximes are more efficiently metabolized in sorghum than in tobacco.

7.5. Conclusion

In conclusion, the current study represents a unique approach which highlights the use of traditional statistical visualization techniques to decipher the biological understanding of oxime metabolism in different systems. From the results, it can be seen that HCA and OPLS-DA SUS-plots are capable of showing the different trends of substrate metabolism which are not easy to interpret from traditional PCA models. Though the results of HCA are not statistically

represented, qualitative observations can be seen and a major conclusion regarding the metabolic patterns within the data can be drawn. On other hand, although the results of the SUS-plot are more qualitative than quantitative, they are statistically profound as the principle of correlation and covariance is taken into consideration during the calculations [31]. Above all, the metabolic tree showed very unique and more statistically important results than the other two models. The results as represented by the HCA and the tree diagrams show that the two plant systems managed to recognize INAP, adapted and metabolized it, and that the biochemical profile is re-adjusting to internal equilibrium over time. It seems that oximes, as seen with the case of INAP, are more efficiently metabolized by cyanogenic as opposed to non-cyanogenic plants. However, the fact that INAP was capable of inducing metabolic perturbations in tobacco as a non-cyanogenic plant is evidence that it is still recognized by the tobacco enzymatic machinery. As shown by our previous work [26], INAP does not only exhibit metabolic-altering capabilities but, at the same time, undergoes biotransformation. These biotransformation events occur by incorporation of certain chemical groups/motifs which are quite similar to those inherited by certain biological molecules with definitive functions towards stress responses [26].

This study extends our knowledge of the metabolism of oximes in plants, especially those that are not cyanogenic. The use of HCA, OPLS-DA-based SUS-plots and metabolic trees in understanding the underlying trend of biological responses at metabolic level is shown here. All these models undisputedly managed to show the time-dependent trend of INAP conversion and associated metabolic changes which are inherit within the metabolomic data generated from two comparable plant systems. Through the use of these models, the results show the response triggered by INAP in sorghum to be more uniform as compared to tobacco where a more variable response was obtained. HCA dendrograms were also shown to be superior to PCA score

plots in representing the results where time-dependent responses are expected. Although quite similar, the metabolic tree diagrams of both tobacco and sorghum were able to reveal differential, underlying clustering patterns similar to those revealed by the HCA dendrograms. It must, however, be stressed that all such models be utilized as parallel approaches since they uncover distinctive underlying trends which complement each other in gaining insight into the biochemical events taking place.

7.7. References

[1]J.K. Nicholson, J.C. Lindon, E. Holmes, 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli *via* multivariate statistical analysis of biological NMRspectroscopic data, Xenobiotica 29 (1998) 1181-1189.

[2]S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast genome, Trends Biotechnol. 16 (1998) 373-378.

[3]H. Theodoridis, G. Gika, I.D. Wilson, Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies, Mass Spectrom. Rev. 30 (2011) 884-906.

[4]C.Y. Lin, M.R. Viant, R.S. Tjeerdema, Metabolomics: methodologies and applications in the environmental sciences, J. Pestic. Sci. 31 (2006) 245-251.

[5]M.R. Viant, Metabolomics of aquatic organisms: the new 'omics' on the block, Mar. Ecol Prog. Ser. 332 (2007) 301-306.

[6]J. Van der Greef, P. Stroobant, R. Van der Heijden, The role of analytical sciences medical systems biology, Curr. Opin. Chem. Biol. 8 (2004) 559-565.

[7]J.B. German, S.M. Watkins, L.B. Fay, Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health, J. Am. Diet Ass. 10 (2005) 1425-1432.

[8]F. Fava, J.A. Lovegrove, R. Gitau, K.G. Jackson, K.M. Tuohy, The gut microbiota and lipid metabolism: implications for human health and coronary heart disease. Curr. Med., Chem. 13 (2006) 3005-3021.

[9]R. Goodacre, Metabolomics of a superorganism, J. Nutr. 137 (2007) 259-266.

[10] J. Kopka, A. Fernie, W. Weckwerth, Y. Gibon, M. Stitt, Metabolite profiling in plant biology: platforms and destinations, Genome Biol. 5 (2004) e109.

[11] W. Weckwerth, K. Morgenthal, Metabolomics: from pattern recognition to biological interpretation, Drug. Discov. Today 10 (2005) 1551-1558.

[12] R.D. Hall, Plant metabolomics: from holistic hope, to hype, to hot topic, New Phytol. 169 (2006) 453-468.

[13] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based metabolomic analysis of plants, Nat. Protoc. 5 (2010) 536-549.

[14] X. Lu, X. Zhao, C. Bai, C. Zhao, G. Lu, G. Xu, LC-MS-based metabonomics analysis, J. Chromatogr. B 866 (2008) 64-76.

[15] I. Olivier, D.T. Loots, A metabolomics approach to characterise and identify various Mycobacterium species, J. Microbiol. Meth. 88 (2012) 419-426.

[16] W.B. Dunn, N.J.C. Bailey, H.E. Johnson, Measuring the metabolome: current analytical technologies, Analyst 130 (2005) 606-625.

[17] J.W. Allwood, R. Goodacre, Introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses, Phytochem. Anal. 21 (2010) 33-47.

[18] R.S. Plumb, I.D. Wilson, High throughput and high sensitivity LC/MS-OA-TOF and UHPLC/TOF-MS for the identification of biomarkers of toxicity and disease using a metabonomics approach, Abstracts of Papers of the American Chemical Society 228 (2004) U164.

[19] M.E. Swartz, UHPLC (TM): an introduction and review, J. Liq. Chromatogr. Relat. Technol. 28 (2005) 1253-1263.

[20] I.D. Wilson, J.K. Nicholson, J. Castro-Perez, J.H. Granger, K.A. Johnson, B.W. Smith, R.S. Plumb, High resolution "Ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies, J. Proteome Res. 4 (2005) 591-598.

[21] D.G. Robertson, Metabonomics in toxicology: a review, Toxicol. Sci. 85 (2005) 809-822.

[22] C.J. Omiecinski, J.P.V. Heuvel, G.H. Perdew, J.M. Peters, Xenobiotic metabolism, disposition, and regulation by receptors: From biochemical phenomenon to predictors of major toxicities, Toxicol. Sci. 120 (2011) 49-75.

[23] J. Pollier, T. Mosesab, A. Goossens, Combinatorial biosynthesis in plants: A review on its potential and future exploitation, Nat. Prod. Rep. 28 (2011) 1897-1916.

[24] I.A. Dubery, C.W. Holzapfel, C.J. Kruger, J.C. Schabort, M. Van Dyk, Characterization of a gamma-radiation induced antifungal stress metabolite in citrus peel, Phytochemistry 27 (1988) 2769-2772.

[25] I.A. Dubery, A.E. Louw, F.R. Van Heerden, Synthesis and evaluation of 4-(3-methyl-2 butenoxy) isonitrosoacetophenone, a radiation-induced stress metabolite in Citrus, Phytochemistry 50 (1999) 983-989.

[26] N.E. Madala, L.A. Piater, P.A. Steenkamp, I.A. Dubery, Biotransformation of isonitrosoacetophenone (2-keto-2-phenyl-acetaldoxime) in tobacco cell suspensions, Biotechnol. Lett. 34 (2012) 1351-1356.

[27] M.T. Werth, S. Halouska, M.D. Shortridge, B. Zhang, R. Powers, Analysis of metabolomic PCA data using tree diagrams, Anal. Biochem. 399, (2010) 58-63.

[28] H. Yamamoto, H. Yamaji, Y. Abe, K. Harada, D. Waluyo, E. Fukusaki, A. Kondo, H. Ohno, H. Fukuda, Chemometrics and intelligent laboratory systems dimensionality reduction for metabolome data using PCA, PLS, OPLS, and RFDA with differential penalties to latent variables, Chemometr. Intell. Lab. Sys. 98 (2009) 136-142.

[29] I.T. Jolliffe, Principal Component Analysis. Second ed. Springer, New York, 2002

[30] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Metabolite profiling for plant functional genomics, Nat. Biotechnol. 18 (2000) 1157-1161.

[31] S. Wiklund, E. Johansson, L. Sjo, J.P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, Visualization of GC / TOF-MS-based metabolomics data for identification of biochemicallly interesting compounds using OPLS class models, Anal. Chem. 80 (2008) 115-122.

[32] M. Steinbach, L. Ertoz, V. Kumar, Challenges of clustering in high dimensional data, University of Minnesota Supercomputing Institute Research Report, 213 (2003) 1-33.

[33] G.M. Downs, J.M. Barnard, Clustering methods and their uses in computational chemistry, in: K.B. Lipkowitz, D.B. Boyd (Eds.) Reviews in Computational Chemistry, Wiley, United Kingdom, 2002, pp. 1-40.

[34] M. Daszykowski, B. Walczak, D.L. Massart, Density-based clustering for exploration of analytical data, Anal. Bioanal. Chem. 380 (2004) 370-372.

[35] J.H. Ward, Hierarchical grouping to optimize an objective function, J. Am. Stat. Ass. 58 (1963) 236-245.

[36] S. Sato, M. Arita, T. Soga, T. Nishioka, M. Tomita, Time-resolved metabolomics reveals metabolic modulation in rice foliage, BMC Syst. Biol. 13 (2008) 1-13.

[37] S. Bak, C.E. Olsen, B.A. Halkier, B.L. Møller, Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and

CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates, Plant Physiol. 123 (2000) 1437-1448.

[38] B.L. Møller, Dynamic metabolons, Science 330 (2010) 1328-1329.

[39] R. Veterinary, K. Word, The biosynthesis of cyanogenic glucosides in roots of cassava, Science 39 (1995) 323-326.

[40] S. Halouska, R.J. Fenton, R.G. Barletta, R. Powers, Predicting the in vivo mechanism of action for drug leads using NMR metabolomics, ACS Chem. Biol. 7 (2012) 166-171.

[41] J. Van der Greef, A.K. Smilde, Symbiosis of chemometrics and metabolomics: past, present, and future, J. Chemometr. 19 (2005) 376-386.

[42] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), J. Chemometr. 16 (2002) 119-128.

[43] B. Bylesjo, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, J. Chemometr. 20 (2006) 341-351.

[44] N.M. Sanabria, I.A. Dubery, Differential display profiling of the *Nicotiana* response to LPS reveals elements of plant basal resistance, Biochem. Biophys. Res. Comm. 344 (2006) 1001-1007

[45] R. Ngara, J. Rees, B.K. Ndimba, Establishment of sorghum cell suspension culture system for proteomics studies, Afr. J. Biotechnol. 7 (2008) 744-749.

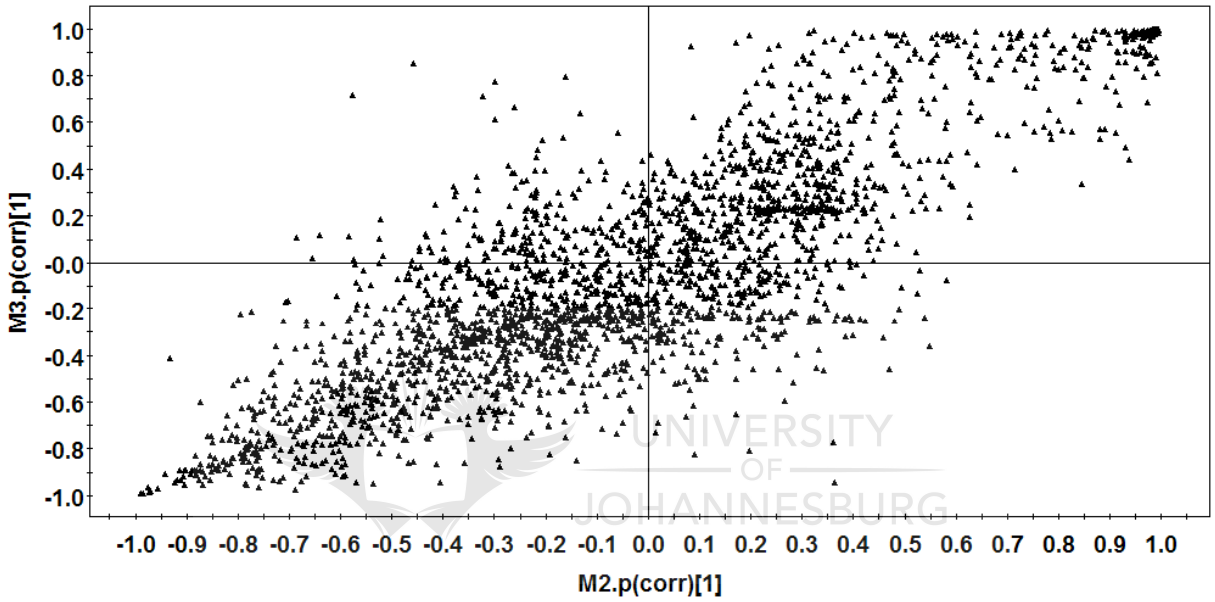
[46] J.D. Retief, Phylogenetic analysis using PHYLIP, Meth. Mol. Biol. 132 (2000) 243-258.

7.6. Supplementary files

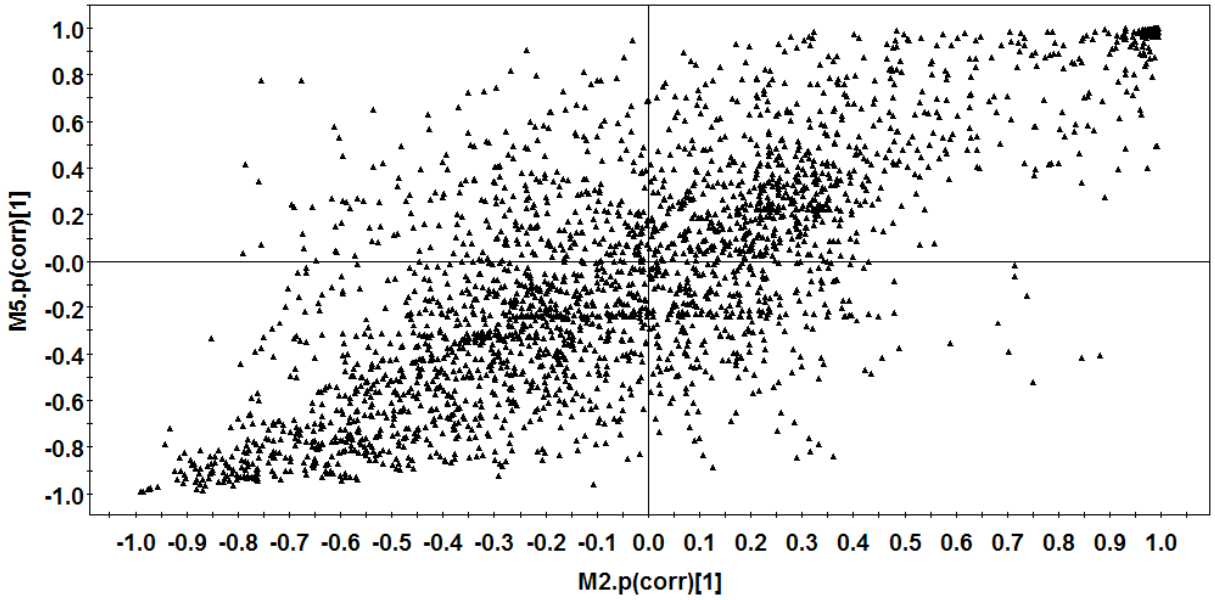
Supplementary Material: Figure S1: OPLS-DA based SUS-plots showing metabolite distribution from different treatment time intervals, **A** (6 h vs12 h), **B** (6 h vs 24 h), **C** (12 h vs 18 h), **D** (12 h vs 24 h) and **E** (18 h vs 24) of **tobacco cell suspensions** treated with 1 mM isonitrosacetophenone. For keys features, refer to Fig. 4.

S1-

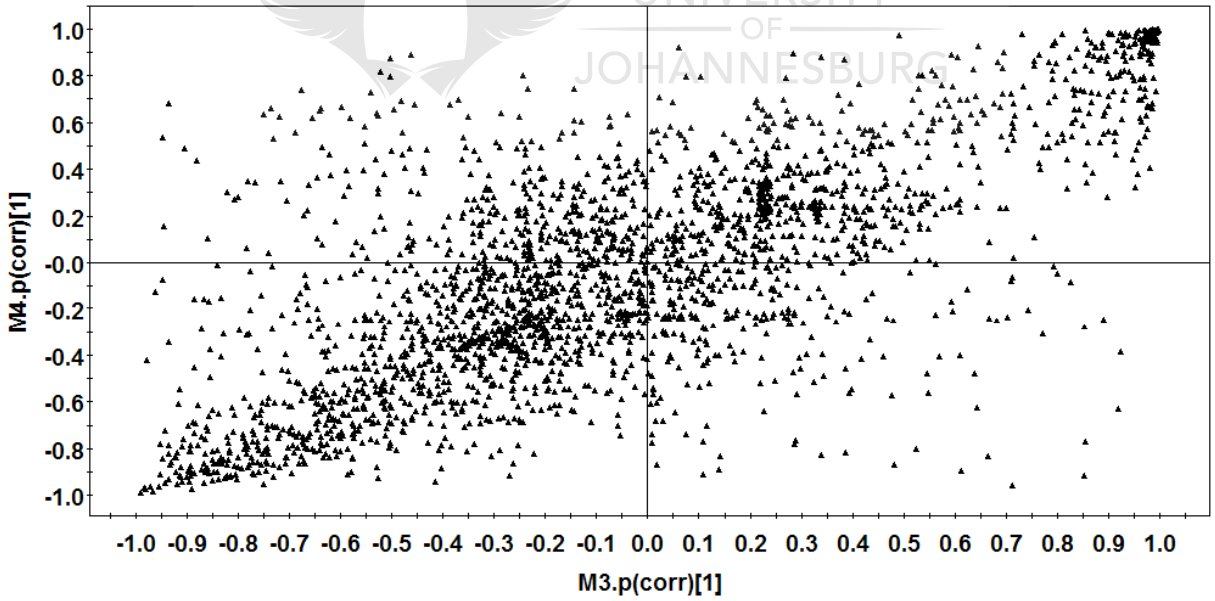
A.



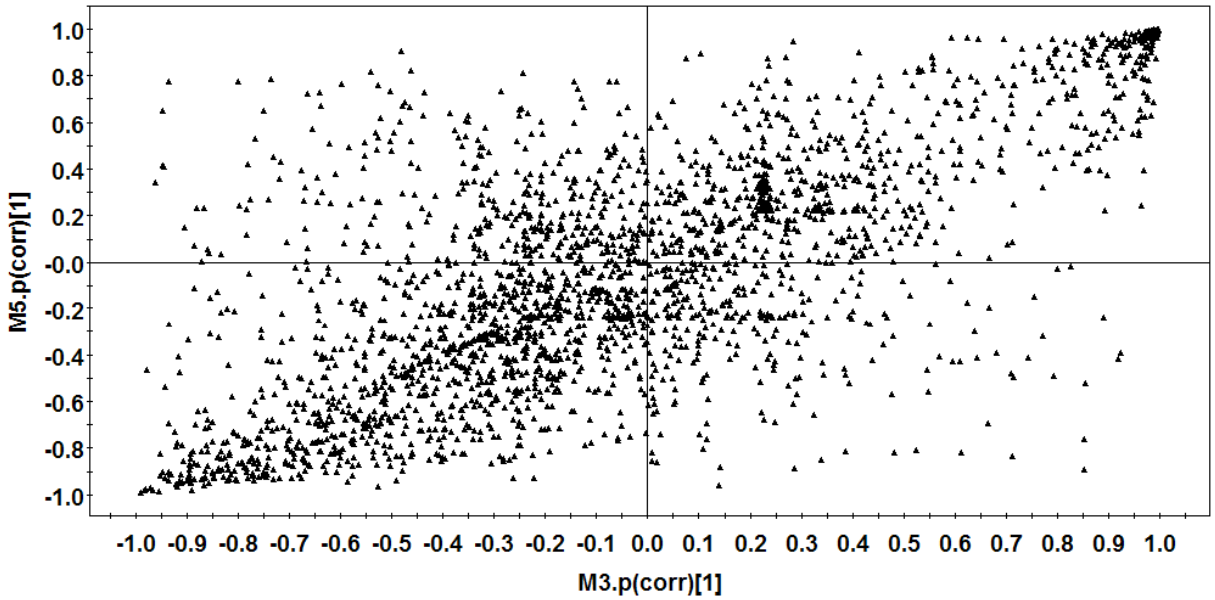
S1-B.



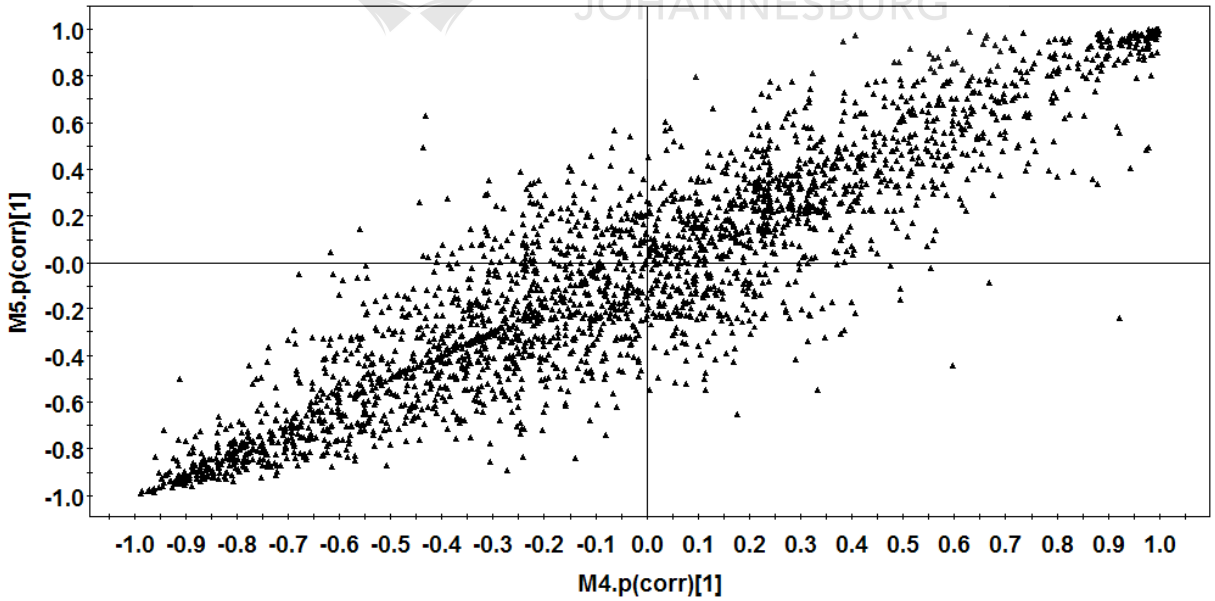
S1-C.



S1-D.

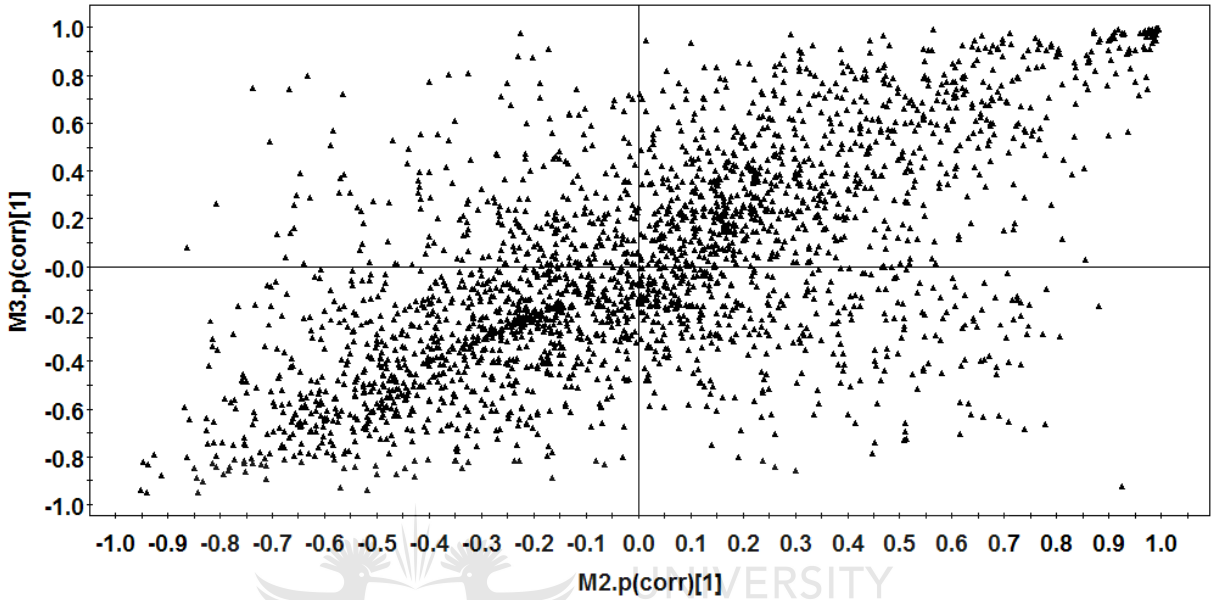


S1-E.

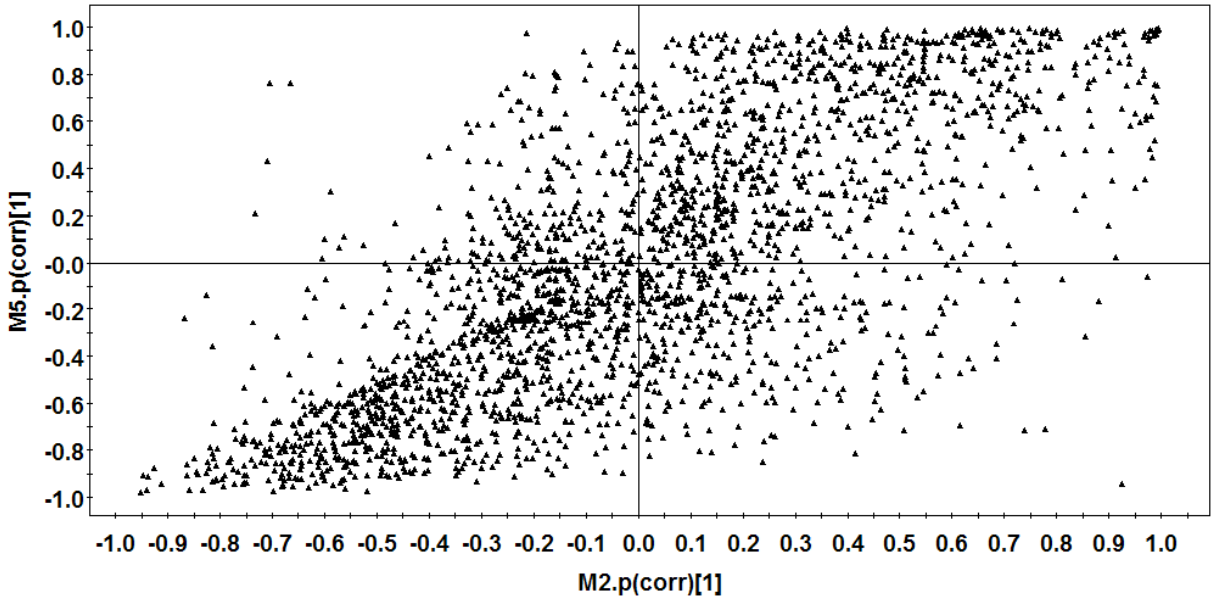


Supplementary Material: Figure S2: OPLS-DA based SUS-plots showing metabolite distribution from different treatment time intervals, **A** (6 h vs.12 h), **B** (6 h vs. 24 h), **C** (12 h vs. 18 h), **D** (12 h vs. 24 h) and **E** (18 h vs. 24) of **sorghum cell suspensions** treated with 1 mM isonitrosacetophenone. For keys features, refer to Fig. 4.

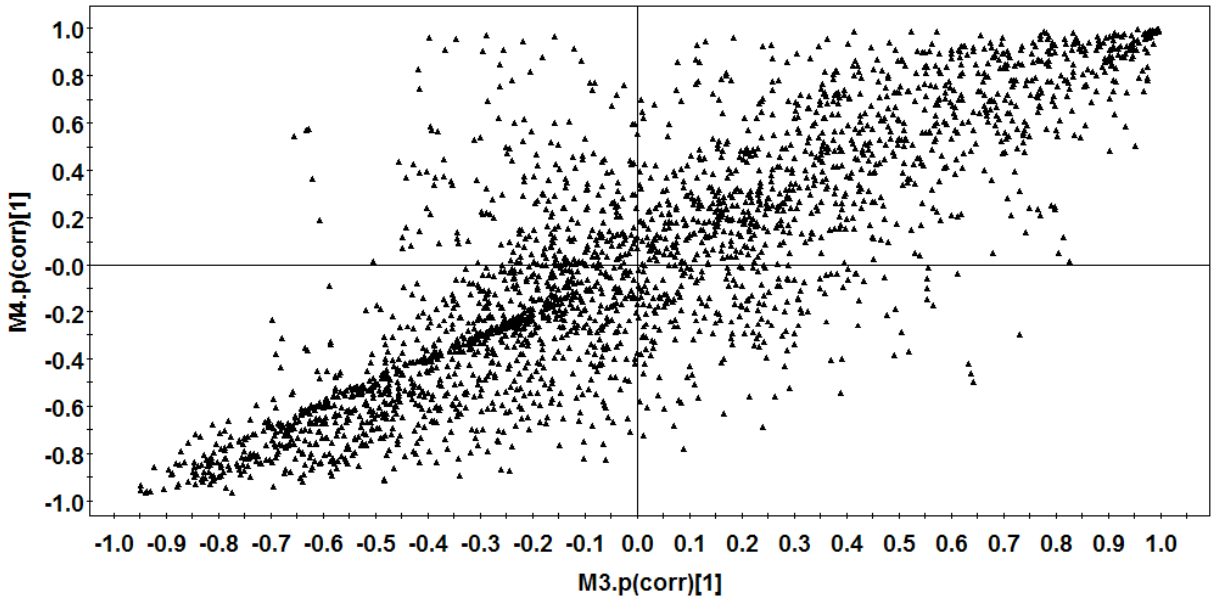
S2-A.



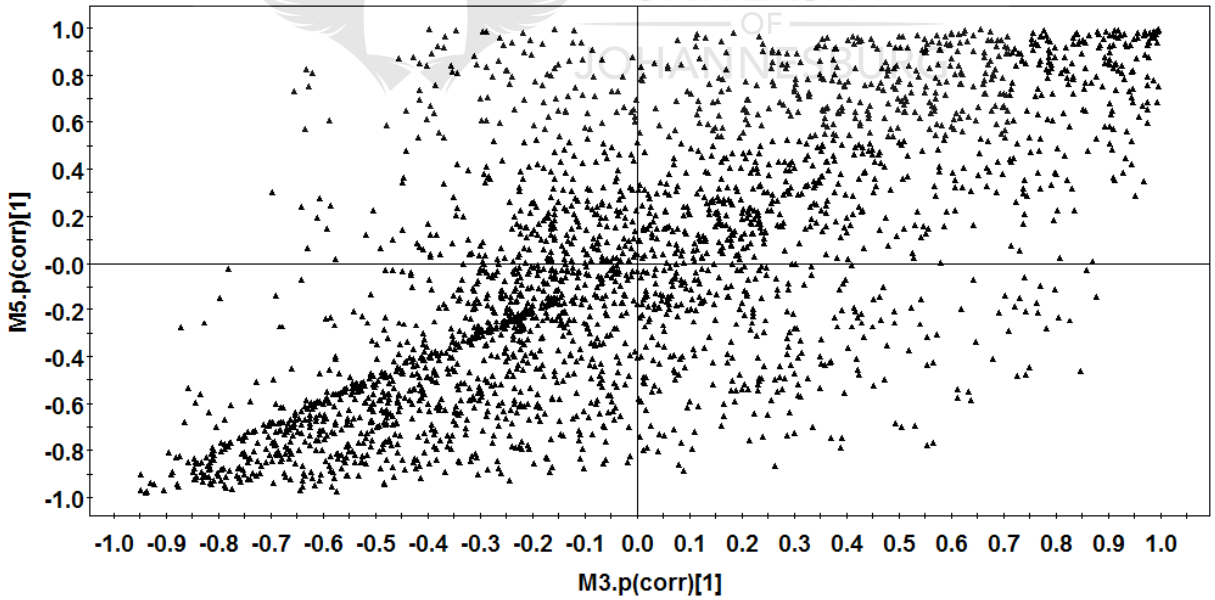
S2-B.



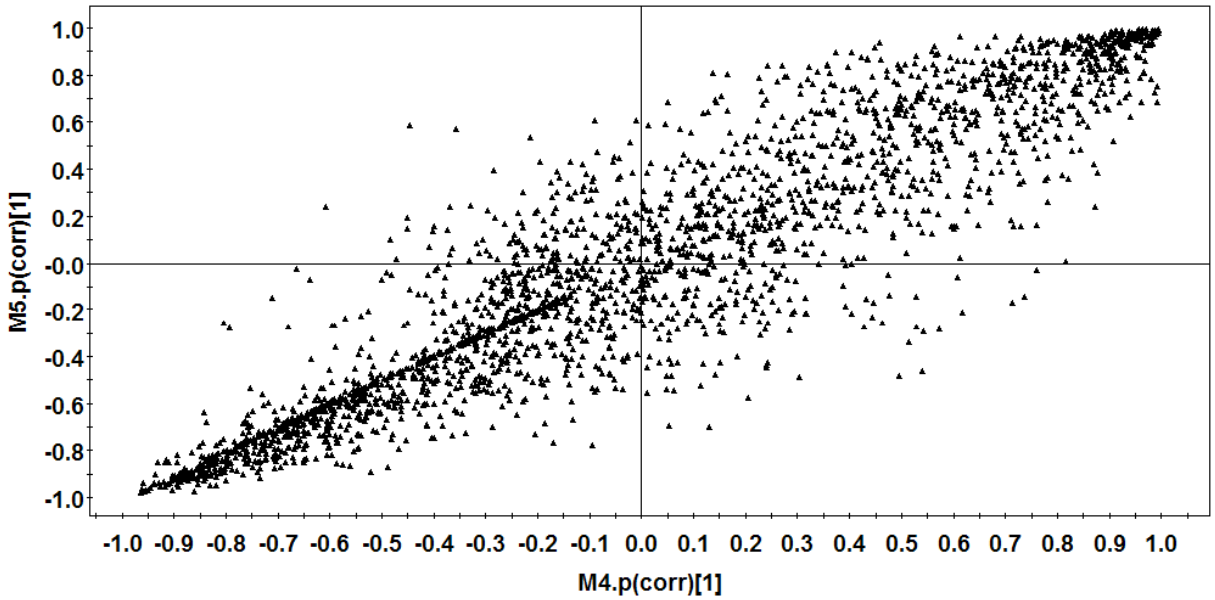
S2-C.



S2-D.



S2-E.



Chapter 8: Biotransformation of isonitrosoacetophenone (2-keto-2-phenyl-acetaldoxime) in tobacco cell suspensions.

8.1. Abstract

Nicotiana tabacum cell suspensions were treated with isonitrosoacetophenone (INAP), a subcomponent of a plant-derived stress metabolite with anti-oxidative and anti-fungal properties. Upon addition of INAP to 1 mM, the cell suspensions exhibited a rapid uptake of INAP and metabolism to a bioconverted product that reached a plateau at 12 h. At 18 h following addition of INAP, 54% of the total intracellular INAP content was represented by the bioconverted product with unconverted INAP at 33.4 µM. UHPLC-MS/MS analyses with MassFragment™ software, were employed for metabolite identification. INAP is subject to biotransformation *via* multiple chemical modifications on the core structure. These include hydroxylation on the meta- and para- positions as well as subsequent methoxylation and glycosylation to result in 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone. The data furthermore propose that INAP is recognized by the enzymatic machinery of the phenylpropanoid pathway and bioconverted to a molecule with a substitution pattern similar to ferulic acid.

Keywords Biotransformation; isonitrosoacetophenone; 2-keto-2-phenyl-acetaldoxime; metabolism; *Nicotiana tabacum;* Solanaceae; xenobiotics.

8.2. Introduction

Biotransformation has been widely exploited as a system of producing novel compounds. Some of the enzymes involved in secondary metabolism may exhibit wide specificities, and the developmentally and spatially controlled expression thereof contributes to tissue and plant-specific chemical phenotypes (Vogt, 2010). Novel enzyme-substrate combinations *in vivo* can lead to the biosynthesis of new, natural product-derived compounds (Pollier et al. 2011).

Citaldoxime, 4-(3-methyl-2-butenoxy)-isonitrosoacetophenone, a novel oxime-containing stress metabolite from citrus, was reported to exhibit phytoalexin and anti-oxidant activities (Dubery et al. 1999). Oxime functional groups are rare in natural products; in plants, oximes are intermediates of a range of metabolic pathways subject to controls that result in variation in both the type and amount of end product formed (Mahandevan, 1973). In the context of plant defence responses, aldoximes are intermediates during the biosynthesis of glucosinolates and cyanogenic glycosides (Møller, 2010). The reactive nitroso-compounds, originating from oxime metabolism, may also act as signals in innate immune responses (Møller, 2010). This interesting hypothesis warrants further investigation into oxime metabolism in plants to elucidate pathways in which such molecules participate.

Metabolites accumulate as the end products of cellular metabolism and the levels thereof reflect the organism's ultimate response to biological or environmental changes (Allwood and Goodacre, 2010). Metabolomics is an unbiased approach aimed at measuring/profiling the metabolite content of the cell, tissue or organism under a specific physiological status and, by studying the metabolome, information regarding the metabolic status and altered pathways can be explained.

In the current study, isonitrosoacetophenone was used as substrate to investigate biotransformation events in tobacco cell suspensions. This report is the first in which the metabolism of INAP in a plant model was investigated.

8.3. Materials and methods

8.3.1. Cell culture, treatment and viability

Nicotiana tabacum cv. Samsun cell suspensions were cultivated as previously described (Sanabria and Dubery, 2006). Three days after sub-culturing, 2 g cells (wet weight) / ml suspension were treated with 250 mM isonitrosoacetophenone (INAP), dissolved in acetone, to a final concentration of 1 mM with continuous rotation at 80 rpm and 25ºC. The final concentration of acetone was 0.4%. Control cells received no treatment. The experimental design included three biological replicates with five technical repeats. Cell viability was evaluated using the TTC method (Towill and Mazur, 1975) and INAP had no toxic effects over 24 h (data not shown).

8.3.2. Extraction and sample preparation

After elicitation time intervals of 6, 12, 18 and 24 h, cells were collected by means of centrifugation at 10 000 x *g* for 10 min in a microcentrifuge at room temperature. Metabolites were extracted from pelleted cells by homogenization in 1 ml 100% methanol. Extraction was further maximised heating the homogenate at 60ºC for 10 min followed by mild sonification for 20 min. The homogenates were then centrifuged at 10 000 x *g* for 10 min, the resulting supernatants transferred to new tubes and dried to completeness using a speed vacuum centrifuge operating with constant heating at 50ºC. The residual brown precipitates were re-dissolved in

400 µl, 50% (v/v) methanol in water. Debris from the samples was removed through a 0.22 µm filter, the filtrates transferred to glass vials fitted with 500 µl inserts and capped.

8.3.3. Ultra-performance liquid chromatography-mass spectrometry

Methanol extracts (5 µl) were analyzed on a Waters UHPLC coupled to a Waters Synapt high definition mass spectrometer (HD-MS) instrument (Waters Corporation, Milford, USA). Chromatographic separation was accomplished using an Acquity BEH C18 column (100 × 2.1 mm with particle size of 1.7 µm) (Waters Corporation) with deionized water (A) and methanol (B), both containing 0.1% formic acid, as mobile phase solvents (described in legend to Fig 1). In order to sufficiently separate the detected target biomarkers, longer chromatographic separations of 30 min were utilised.

The separated analytes were monitored using both photo-diode array (PDA, 220 - 500 nm) and electro-spray ionisation time-of-flight mass spectrometry (ESI-TOF-MS) detectors. For MS detection, experimental conditions are described in the legend to Fig. 2. Mass accuracy of the *m/z* values in all the acquired spectra were automatically corrected during acquisition based on the instrumental calibration files and a simultaneous lockmass signal. Data was collected in extended dynamic range mode. The molecular formula assignments were obtained with the MassLynx^{TM} software (version 4.1 SCN 704, Waters Corporation). The area/intensities of the peaks across the whole chromatogram were evaluated/calculated with the aid of the data mining technique, MakerLynx^{TM} (Waters Corporation). Data was further analysed by Minitab, version 15 statistical software.

8.3.4. Quantification

From UHPLC-PDA data, the peaks representing INAP and its biotransformed product were manually integrated and quantified using an INAP standard curve (0.125 - 2.0 mM). Single wavelength (265 nm) monitoring aimed at maximum detection of the two metabolites with application of the Savitzky-Golay smoothing/filtering procedure was used. The areas of these two peaks were calculated with MassLynx^{TM} software, and the % conversion of INAP inside the cells calculated using five independent data sets from INAP treated samples.

8.4. Results and discussion

The biological roles of oxime compounds such as citaldoxime in plants are not well researched with regard to the biosynthesis, interconversions and biological effects. Here, we report on the biotransformation of INAP by the metabolic machinery of *Nicotiana tabacum,* a non-cyanogenic plant. As a central analytical technique in metabolomics, MS provides a combination of rapid, sensitive and selective qualitative and quantitative analyses with the ability to identify metabolites. UHPLC-HD-MS (Allwood and Goodacre, 2010) was used for investigating the metabolites resulting from the bioconversion. MassFragment^{TM} software (Waters Corporation), a computer-based method which uses chemically intelligent algorithms to evaluate proposed structures for observed fragment ions on MS data of small molecule compounds (www.waters.com), was also used to validate the proposed INAP-derived compound. This combination provides a powerful approach to investigating biotransformation events.

Table 1 Time-dependent changes in the INAP content in cells reflecting uptake and conversion to 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone. Cell suspensions (2 g/ml) were initially treated with 1 mM INAP.

Time (h)	$[INAP]_{cell}$ (µM ± SD)	Conversion (%)
0	-	-
6	112.3 ± 1.8	20
12	106.5 ± 8.6	32
18	33.4 ± 4.2	54
24	45.5 ± 9.6	50

8.4.1. Metabolite profiling

The introduction of INAP (**I**) affected the metabolome of tobacco cells (2 g/ml suspension containing 1 mM INAP), and time-dependent changes in the metabolite profiles were indicated by the UHPLC chromatograms (Fig. 1). Three major peaks were found to exist in the INAP-treated samples as compared to the controls, one of which is the biotransformed product (**II**, 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone). The peak corresponding to (**I**) showed a time-dependent decrease from 6 h onwards, concomitant with an increase in (**II**), indicating metabolism and bioconversion of INAP. As shown in Table 1 and the distribution graph of (**II**) (Fig. 2), the biotransformation increased from 6 h onwards and reached a plateau at 12 h.

Quantification of residual cellular INAP at 18 h revealed that 33.4 µM remained unconverted with 54% of the total intracellular INAP represented by (**II**).

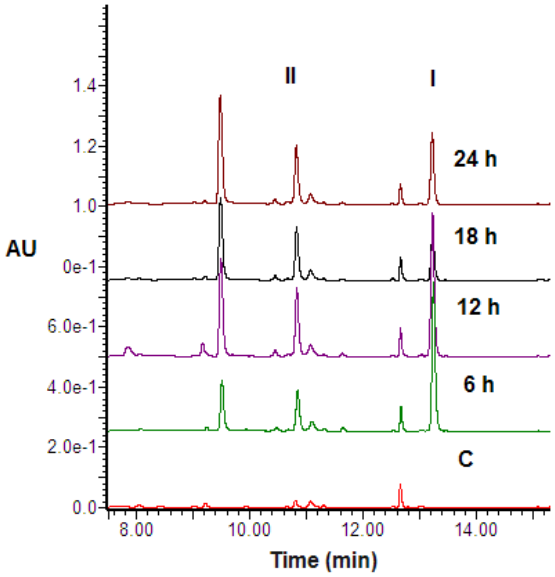


Fig. 1 Comparison of representative UHPLC-PDA base peak intensity (BPI) chromatograms of extracts from tobacco cell suspensions treated with INAP for 0, 6, 12, 18 and 24 h. The composition of mobile phase A was 0.1% formic acid in deionized water and mobile phase B was 0.1% formic acid in methanol. The column was eluted with a linear gradient at a constant flow rate of 400 µl/min of 5% B over 0.0-2.0 min, 5-95% B over 2.0-22.0 min, held constant at 95% B over 22.0-25.0 min, 95-5% B over 25.0-27.0 min and a final wash at 5% B over 27-30 min. The peaks corresponding to INAP and the biotransformed product are indicated by (**I**) and (**II**) respectively.

Although a structurally similar compound, 4-(3-methyl-2-butenoxy)-isonitrosoacetophenone was reported as a stress metabolite in citrus (*Citrus sinensis*, Rutaceae) (Dubery et al. 1999), INAP can be regarded as a xenobiotic compound that is metabolized by the metabolic machinery present in tobacco cells. Enzymes involved in metabolism of secondary

metabolites often exhibit limited specificity (Schwab, 2003). These multifunctional, but poor substrate-specific, enzymes may also contribute to metabolome diversity through the capability of accepting more than one substrate, thereby catalyzing more than one reaction and forming multiple products (Vogt, 2010). In plants, like in microorganisms, many biosynthetic pathways are not fully active, or at least not detectable at the resolution of the present metabolite profiling technologies. This 'silent metabolism' (Lewinsohn and Gijzen, 2009) can be activated by either providing the appropriate substrate or by triggering the inactive parts of the biosynthetic pathways. Biotransformation can also fulfil the function of eliminating xenobiotic compounds from the system (Omiecinski et al. 2011) but this can alter the physical properties of a compound, for instance glycosylation of flavonoids increases the water solubility and enhances bioavailability of the compounds. Results obtained in this study suggest that INAP undergoes a series of modifications characterized by addition of different chemical groups on its ring structure.

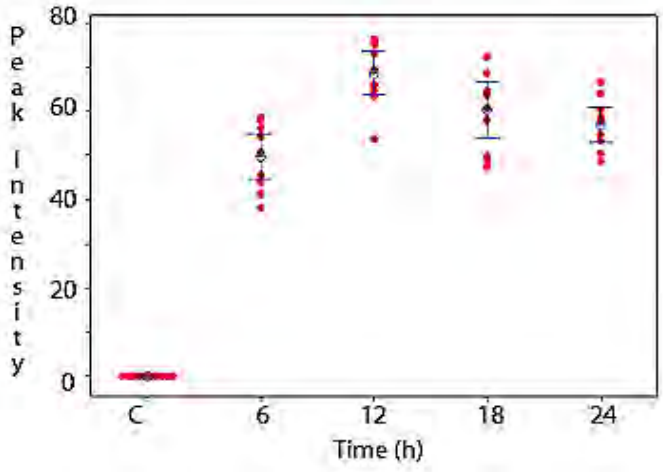


Fig. 2 Distribution graph showing the dispersal of the *m/z* 356.0964 ion (Fig. S1, S2) across the INAP-treated samples at different time intervals in comparison to the untreated control. The

target represents the mean and the bars represent the region of 95% confidence interval of the mean. Experimental conditions: ESI^- , capillary voltage: 2.5 kV, sample cone voltage: 17 V, extraction cone voltage: 5.0 V, MCP detector voltage: 1600 V, source temperature: 120°C, desolvation temperature: 350°C, cone gas flow: 50 (l/h), desolvation gas flow: 450 (l/h), *m/z* range: 100-1000, scan time: 0.1 sec, interscan delay: 0.02 sec, mode: centroid, lockmass: leucine enkephalin (556.3 g/mol), lockmass flow rate: 0.4 ml/min, mass accuracy window: 0.5 Da.

8.4.2. Structural modifications to the INAP core structure

From the ESI spectrum, the biotransformed product was detected as *m/z* 356.0964 and using this mass, the accurate mass of (**II**) was further verified with the aid of MS/MS experiments. The results of this analysis revealed only one product exclusively in treated samples (Fig. S1). Using the data from MarkerLynx™ analysis, the distribution graph showing the levels of (**II**) across the different time intervals in comparison to the controls was also constructed (Fig. 2). MassFragment™ software was used to validate the proposed structure, and to determine whether it correlates with the fragmentation pattern as seen on the MS/MS spectral data (Fig. S2). The MassFragment™ output (Table S1) shows that some of the fragment ions comply with possible fragmentation sites of the proposed biotransformed metabolite **II** (Fig. 3). From the proposed INAP biotransformation pathway (Fig. 3) it is thus conceivable that INAP is hydroxylated on the 3' and 4' positions before being methoxylated on the 3' position by an *O*-methyl transferase (*O*-MT) to generate a substitution pattern on the aromatic ring similar to that of ferulic acid. The addition of a glucose residue by an *O*-glucosylase would then render compound **II,** 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone. These proposed structural modifications are all supported by the observed ESI^- fragmentation patterns.

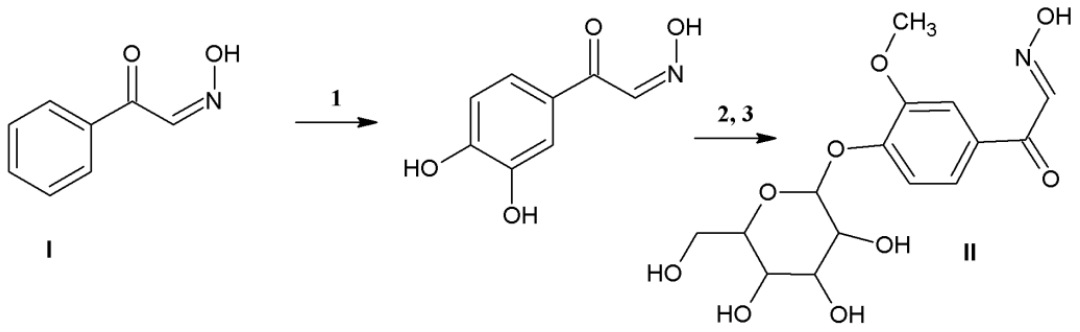


Fig. 3 The chemical structure of isonitrosoacetophenone (**I**) and the biotransformed product (**II**), 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone. The proposed biotransformation events of INAP in tobacco cell suspensions are 1 = hydroxylation on the *m* and *p* positions; 2, 3 = methoxylation and glucosylation. The order of substitution is not certain.

Though no enzymatic investigations were done during the course of this study, cinnamate-4-hydroxylase (C4H), an enzyme found across a wide spectrum of plants, has the ability to metabolize some substrate analogues including xenobiotics (Schalk et al. 1997). These analogues are derivatives of cinnamate with a planar structure, negatively charged side chain and a size of one aromatic ring (Schoch et al. 2003). The metabolism of unnatural substrates by C4H provides the possibility to synthesize novel products from the phenylpropanoid- and benzoic acid pathways.

Furthermore, *O*-MTs have also been shown to act on various natural products like simple catechols, phenylpropanoids and structurally complex isoquinoline alkaloids (Frick and Kutchan, 1999). Tobacco cells possess an array of *O*-MT isoforms with variable efficiency toward the diverse plant ortho-diphenolic substrates (Maury et al. 1999), and stress-induced multifunctional *O*-MTs from *Pinus sylvestris* have been shown to catalyze the methylation of diverse phenolic substrates (Chiron et al. 2000). The very broad substrate specificity included molecules derived from several branches of the phenylpropanoid pathway. In addition to enzymes exhibiting the

ability to accept a wide range of substrates, substrate availability was also shown to contribute to diversity in the formation of the methylated products.

Glucosylated hydroxycinnamic acids, where the glucose is attached to the phenolic hydroxyl group, are often found as intermediates in phenylpropanoid metabolism. In addition, glucosylation of hydroxycinnamic acid derivatives have been suggested as a detoxification mechanism employed in plants against reactive phenylpropanoids (Meyermans et al. 2000).

8.5. Conclusion

INAP can be regarded as a xenobiotic chemical which is recognised by enzymes involved in secondary plant metabolism and subsequently biotransformed to a stable molecule, 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone. INAP was found to be metabolized through a series of steps which involves hydroxylation and subsequent methoxylation and glucosylation. These structural modifications could be the way in which tobacco cell suspensions eliminate INAP from the system by enzymes belonging to secondary metabolic pathways, acting upon precursor molecules with structural similarities to INAP. From the proposed structure of the biotransformed product **II**, this molecule exhibits features/modifications resembling those originating from the phenylpropanoid pathway. For example, ferulic acid glucoside is derived from *trans*-cinnamic acid through similar modifications proposed for INAP metabolism. Ferulic acid and derivatives thereof are effective anti-oxidants (Kikuzaki et al. 2002). The function of 4'-hexopyranosyloxy-3'-methoxyisonitrosoacetophenone as a potential phytoprotective agent is still to be elucidated and warrants further investigation.

8.6. References

Allwood JW, Goodacre R (2010) Introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. Phytochem Anal 21: 33-47

Chiron H, Drouet A, Claudot A-C, et al. (2000) Molecular cloning and functional expression of a stress-induced multifunctional *O*-methyltransferase with pinosylvin methyltransferase activity from Scots pine (*Pinus sylvestris* L.). Plant Mol Biol 44: 733–745

Dubery IA, Louw AE, Van Heerden FR (1999) Synthesis and evaluation of 4-(3-methyl-2 butenoxy) isonitrosoacetophenone, a radiation-induced stress metabolite in Citrus. Phytochemistry 50: 983-989

Frick S, Kutchan TM (1999) Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. Plant J 17: 329-339

Kikuzaki H, Hisamoto M, Hirose K, et al. (2002) Antioxidant properties of ferulic acid and its related compounds. J Agric Food Chem 50: 2161-2168

Lewinsohn E, Gijzen M (2009) Phytochemical diversity: The sounds of silent metabolism. Plant Sci 176: 161-169

Mahadevan S (1973) Role of oximes in nitrogen metabolism in plants. Annu Rev Plant Physiol 24: 69-88

Maury S, Geoffroy P, Legrand M (1999) Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-*O*-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-*O*-methyltransferase classes have distinct substrate specificities and expression patterns. Plant Physiol 121: 215-223

Meyermans H, Morreel K, Lapierre C, et al. (2000) Modifications in lignin and accumulation of phenolic glucosides in Poplar xylem upon down-regulation of caffeoyl-coenzyme A *O*-methyltransferase, an enzyme involved in lignin biosynthesis. J Biol Chem 275: 36899 – 36909

Møller BL (2010) Dynamic metabolons. Science 330: 1328-1329

Omiecinski CJ, Heuvel JPV, Perdew GH, Peters JM (2011) Xenobiotic metabolism, disposition, and regulation by receptors: From biochemical phenomenon to predictors of major toxicities. Toxicol Sci 120:49-75

Pollier J, Mosesab T, Goossens A (2011) Combinatorial biosynthesis in plants: A review on its potential and future exploitation. Nat Prod Rep 28: 1897-1916

Sanabria NM, Dubery IA (2006) Differential display profiling of the *Nicotiana* response to LPS reveals elements of plant basal resistance. Biochem Biophys Res Comm 344: 1001-1007

Schalk M, Pierrel MA, Zimmerlin A, Batard Y, Durst F, Werck-Reichhart D. (1997) Xenobiotics: substrates and inhibitors of the plant cytochrome P450. Environ Sci Pollut Res 4: 229–234

Schwab W (2003) Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62: 837-847

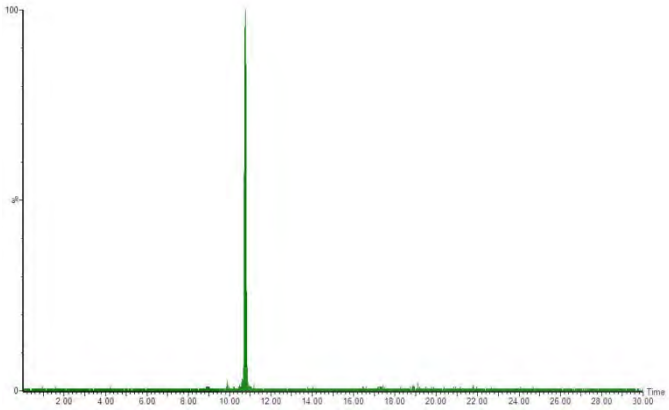
Schoch GA, Attias R, Le Ret M, Werck-Reichhart D (2003) Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1 – homology model guided site-directed mutagenesis. Eur J Biochem 270: 3684–3695

Towill, LE, Mazur, P (1975) Studies on the reduction of 2,3,5-triphenyltetrazolium chloride as a viability assay for plant tissue cultures. Can J Bot 53: 1097-1102

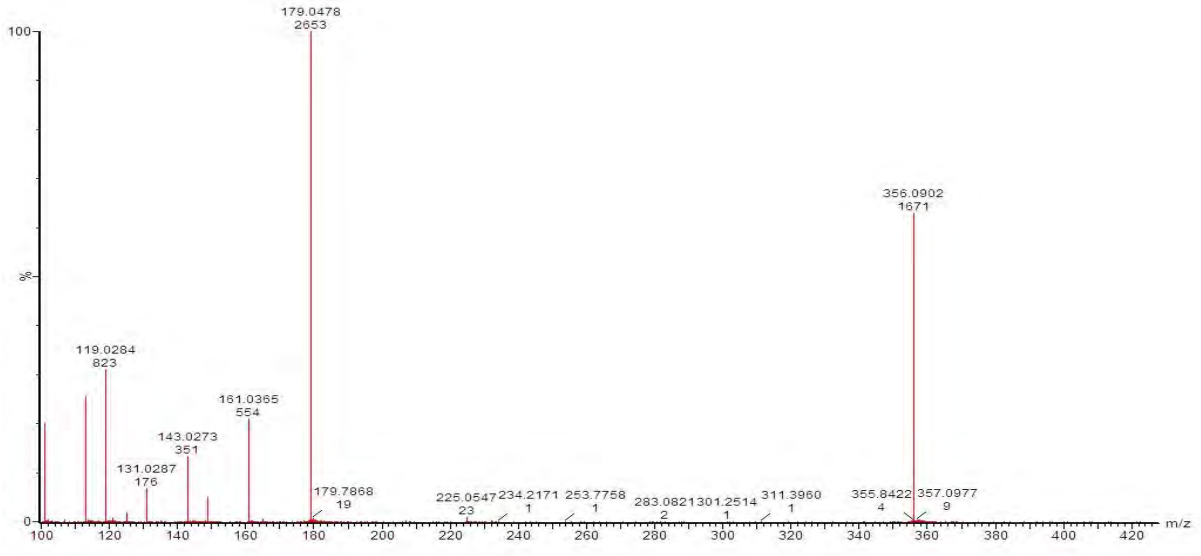
Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant 3: 2-20.

8.7. Supplementary files

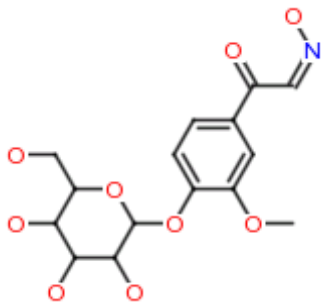
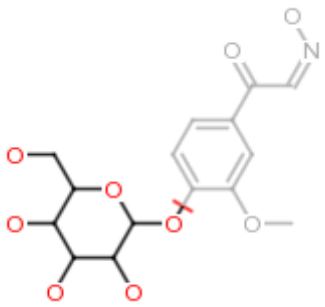
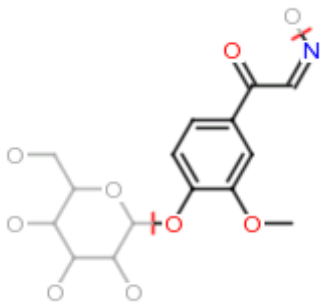
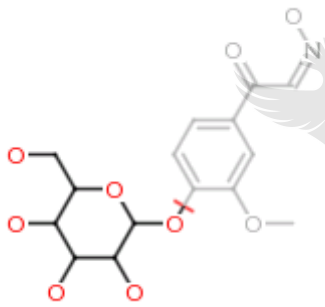
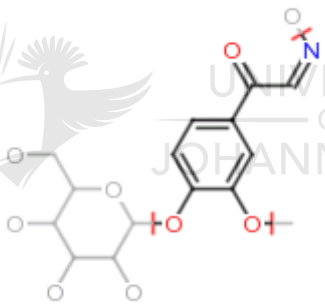
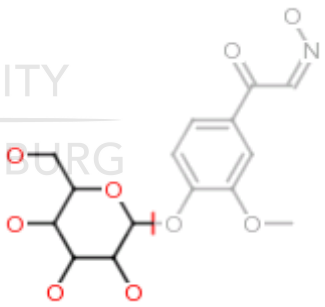
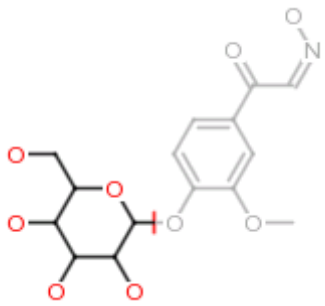
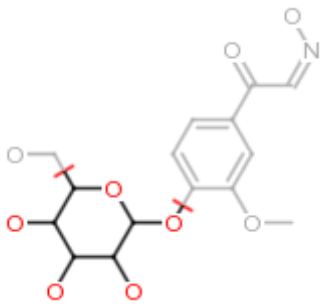
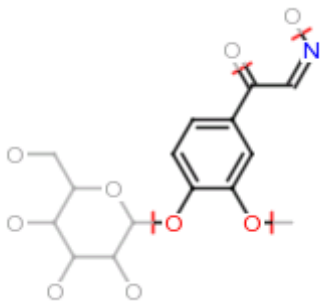
Supplementary Fig. 1 UHPLC-MS/MS BPI chromatogram showing a single peak at Rt of 10.06 min of *m/z* 356.0964 biomarker.

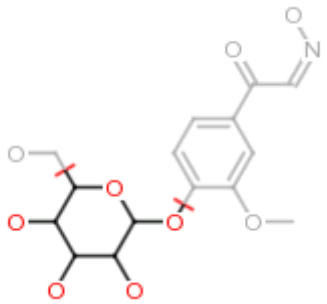
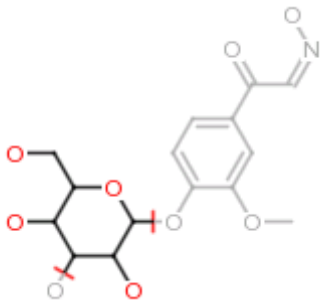
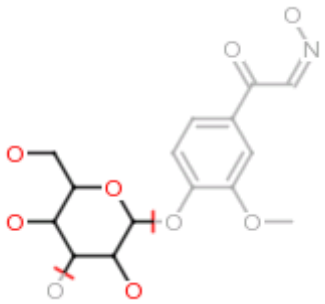
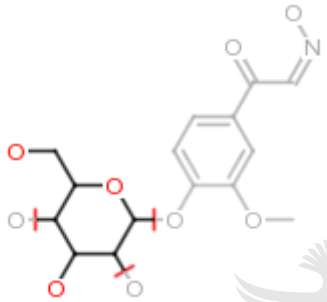
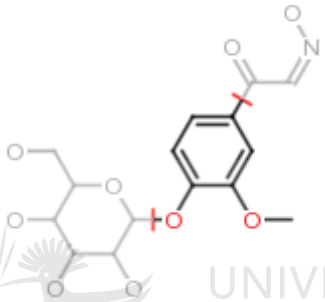
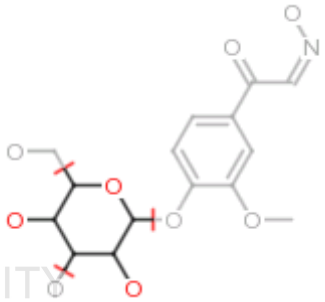


Supplementary Fig. 2 UHPLC-MS/MS spectrum showing the fragmentation patterns of *m/z* 356.0964 ion on ESI negative ionization mode.



Supplementary Table1: MassFragment results, showing the correlation between fragmentation patterns and structural modifications of the m/z = 356.09 ion on ESI negative ionization.

356.0964 ¬- (+1H)  356.0982 (-1.8.mDa) (S:1.0, B:0) $C_{15}H_{18}NO_9$ (-none)	**180.0586** ¬˙- (+3H)  180.0634 (-4.8.mDa) (S:10.0, B:1) $C_6H_{12}O_6$ (-$C_9H_6NO_3$)	**180.0586** ¬- (+5H)  180.0661 (-7.5.mDa) (S:1.5, B:2) $C_9H_{10}NO_3$ (-$C_6H_8O_6$)
179.0533 ¬- (+2H)  179.0556 (-2.3.mDa) (S:10.0, B:1) $C_6H_{11}O_6$ (-$C_9H_7NO_3$)	**165.0421** ¬˙- (+5H)  165.0426 (-0.5.mDa) (S:2.0, B:3) $C_8H_7NO_3$ (-$C_7H_{11}O_6$)	**162.0481** ¬˙- (+1H)  162.0528 (-4.7.mDa) (S:0.5, B:1) $C_6H_{10}O_5$ (-$C_9H_8NO_4$)
161.0440 ¬- (+0H)  161.0450 (-1.0.mDa) (S:0.5, B:1) $C_6H_9O_5$ (-$C_9H_9NO_4$)	**150.0460** ¬˙- (+4H)  150.0528 (-6.8.mDa) (S:11.0, B:2) $C_5H_{10}O_5$ (-$C_{10}H_8NO_4$)	**149.0454** ¬˙- (+5H)  149.0477 (-2.3.mDa) (S:4.0, B:4) $C_8H_7NO_2$ (-$C_7H_{11}O_7$)

148.0437 ˙- (+2H)  148.0372 (+6.5.mDa) (S:11.0, B:2) $C_5H_8O_5$ (-$C_{10}H_{10}NO_4$)	**144.0379** ˙- (+0H)  144.0423 (-4.4.mDa) (S:1.0, B:2) $C_6H_8O_4$ (-$C_9H_{10}NO_5$)	**143.0340** ¬- (-1H)  143.0344 (-0.4.mDa) (S:1.0, B:2) $C_6H_7O_4$ (-$C_9H_{11}NO_5$)
125.0239 ¬- (-2H)  125.0239 (+0.0.mDa) (S:1.5, B:3) $C_6H_5O_3$ (-$C_9H_{13}NO_6$)	**121.0300** ¬- (+1H)  121.0290 (+1.0.mDa) (S:10.5, B:2) $C_7H_5O_2$ (-$C_8H_{13}NO_7$)	**116.0464** ¬˙- (+3H)  116.0473 (-0.9.mDa) (S:2.0, B:3) $C_5H_8O_3$ (-$C_{10}H_{10}NO_6$)

Chapter 9: General conclusion

Oximes are rare functional groups within plant natural products and, as such, the biology and chemistry thereof is not fully understood. However, the occurrence is known in certain plants capable of producing specialized stress-related molecules known as glucosinolates and cyanogenic glycosides. Apart from participation as intermediates during biosynthesis of these compounds, there is no other well-defined function(s) of oximes in plants. The existence in glucosinolates or cyanogenic glycosides non-producing plants have also been reported previously, but without comprehensive scientific explanation. It was therefore the aim of the current study to investigate other possible function of oximes in plants, especially in the afore-mentioned non-producers. Here, INAP, an oxime molecule with structural similarities to citaldoxime, a stress-induced metabolite found in citrus, was used. With the aid of UHPLC-MS based-metabolomics, multivariate data models, and other chemical intelligent software, the effect of INAP on three different plant systems was investigated. INAP was found to trigger metabolite changes in all three systems and to undergo biotransformation events. With the aid of novel metabolite identification procedures, this oxime molecule was found to induce metabolites with known functions toward stress-related responses. From the literature, INAP-induced metabolites are known to exhibit antimicrobial and antioxidant activities. These findings explain the initial accumulation of citaldoxime in citrus peel undergoing oxidative stress due to ionizing radiation treatment. Furthermore, INAP was shown to be more efficiently metabolized in a cyanogenic plant system like sorghum than in non-cyanogenic plants, suggesting evolutionary differences between the three systems for oxime utilization. As stipulated above, INAP was found to undergo chemical modification (biotransformation) events on its core structure, similar to those on known endogenous metabolites with stress-related activities. The results of the current study highlight the feasibility of using UHPLC-MS and chemometric-based models for comprehensive biological understanding of chemical-induced metabolite perturbations and the metabolic fate thereof.