Modern Plant Metabolomics: Advanced Natural Product Gene Discoveries, Improved Technologies, and Future Prospects

Journal:	*Natural Product Reports*
Manuscript ID:	NP-REV-05-2014-000072.R1
Article Type:	Review Article
Date Submitted by the Author:	24-Sep-2014
Complete List of Authors:	Sumner , Lloyd W. ; The Samuel Roberts noble foundation , Plant Biology Division Lei, Zhentian; The Samuel Roberts Noble Foundation, Plant Biology Nikolau, Basil; Iowa State University, Biochemistry, Biophysics and Molecular Biology; U.S. Department of Energy, Ames Laboratory Saito, Kazuki; RIKEN Center for Sustainable Resource Science, ; Chiba University, Graduate School of Pharmaceutical Sciences

Journal Name

RSCPublishing

ARTICLE

Modern Plant Metabolomics: Advanced Natural Product Gene Discoveries, Improved Technologies, and Future Prospects

Cite this: DOI: 10.1039/x0xx00000x

Lloyd W. Sumner [a], Zhentian Lei [a], Basil J. Nikolau[b,c] and Kazuki Saito[d,e]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Plant metabolomics has matured and modern plant metabolomics has accelerated gene discoveries and the elucidation of a variety of plant natural product biosynthetic pathways. This review highlights specific examples of the discovery and characterization of novel genes and enzymes associated with the biosynthesis of natural products such as flavonoids, glucosinolates, terpenoids, and alkaloids. Additional examples of the integration of metabolomics with genome-based functional characterizations of plant natural products that are important to modern pharmaceutical technology are also reviewed. This article also provides a substantial review of recent technical advances in mass spectrometry imaging, nuclear magnetic resonance imaging, integrated LC-MS-SPE-NMR for metabolite identifications, and x-ray crystallography of microgram quantities for structural determinations. The review closes with a discussion on the future prospects of metabolomics related to crop species and herbal medicine.

1. Introduction

Metabolite profiling of bacterial, mammalian, and plant metabolites has been an integral component of biological studies since the 1960s. The continual refinement, increasing scope and larger-scale of plant metabolite profiling have led to the evolution of modern plant metabolomics, which has matured as a valuable tool for advancing our understanding of plant biology and physiology. This review is not intended to be a comprehensive review of plant metabolomics, but instead will focus on some of the more recent advances that continue to propel the utility of plant metabolomics for the discovery and understanding of plant natural products or specialized metabolites. This includes examples of integrated metabolomics and genome-based functional characterizations of plant natural products that are important to modern pharmaceutical technology. The review further highlights a large number of specific examples illustrating the utility of metabolomics for the discovery and characterization of novel genes and enzymes responsible for the biosynthesis of natural products such as flavonoids, glucosinolates, terpenoids, and alkaloids. The review then summarizes recent technical advances in mass spectrometry imaging, nuclear magnetic resonance imaging, integrated LC-MS-SPE-NMR for metabolite identifications, and x-ray crystallography of microgram quantities for structural determinations. The review closes with a summary on potential opportunities for metabolomics related to crops and herbal medicines.

2. Genomics and metabolomics mining of medicinal plants

Medicinal plants have a long association with human history, being part of our folklore and the basis of natural medicines that have been discoursed under such topics as herbalism, phytotherapy, ethnobotany and ethnopharmacy. History is replete with examples of plants being used to affect the human condition; maybe a famous example being that of hemlock, the fatal poison extract that the condemned Socrates was forced to drink.[1] However, archaeological studies appear to provide evidence of medicinal plant use some 50,000 years ago in the Paleolithic age, and written evidence can be traced to the Sumerians, about 5,000 years ago.[2] At the start of the 19th century, evidence based medical practices began to replace such traditional or alternative medicines, but the use of medicinal plants is still almost universal among non-industrialized societies,[3] and plants still serve as the source of many nutraceuticals and pharmaceuticals.

Plant-sourced biochemicals are an important part of our modern pharmaceutical technology. Examples of such current medicines with a long history are aspirin, the acetyl-ester of salicylic acid that was originally isolated from the bark of the willow tree;[4] and the antimalarial alkaloid, quinine isolated in the 19th century from the bark of trees of the cinchona genus.[5] More recent natural product based pharmaceuticals include paclitaxel (more commonly known as taxol™), isolated from the bark of the Pacific yew tree, and used in cancer chemotherapy,[6, 7] and artemisinin, isolated from sweet wormwood (*Artemisia annua*), and used as an anti-malarial drug.[8] Such bioactive compounds have been traditionally identified and characterized following the fractionation and purification of extracts, guided by bioactivity assays, a strategy that has been at the core of the field of natural product chemistry.[9, 10] Many of these bioactive compounds are products of secondary or specialized metabolism, and as such their taxonomic distribution is in relatively narrow phylogenetic clades within *Plantae*. Because of the narrow phylogenetic distribution of these phytochemicals, the biochemical

and genetic components that support their biosynthesis cannot be identified and characterized by homology to a model system, but have to be investigated directly by studying the organism that produces that phytochemical. The role of metabolomics in such studies is two-fold: 1) to identify the spatial and temporal distribution of the target phytochemical as influenced by plant development and environmental cues; and 2) identify related compounds, which may be considered as either intermediates of biosynthesis or alternative products of promiscuous enzymes that support the biosynthesis of the target phytochemical. Such data, when integrated with the generalizable principles of organic synthetic chemistry, can provide avenues to discovering biosynthetic pathways. Integrating metabolomics therefore with genome-based functional characterizations of gene products is providing an accelerated path to discovering novel biosynthetic pathways to specialized metabolites.

With the advent of advanced genomics capabilities, associated with the ability to comprehensively and quickly determine and assemble the genomes of plants, there has been a logical need to understand how this genetic information is expressed in a comprehensive manner. This has led to the emergence of such fields as transcriptomics, proteomics and metabolomics. Although technical difficulties have been associated with implementing these global genome expression-profiling technologies, advances in nucleic acid sequencing has enabled the acquisition of near comprehensive transcriptomic datasets, for even the most complex genetic organisms. Moreover, as a consequence of the Central Dogma, and the near linear relationship between the genome, transcriptome and proteome, one is able to predicate the primary structures of these three levels of genetic expression from any one other. However, it is not possible to predict the chemical nature of the metabolome from genetic-based datasets because the relationship between the proteome and metabolome is not linear but redundant. Despite this limitation, the past decade has seen considerable efforts and successes in the integration of genomics and metabolomics datasets to uncover new knowledge concerning the biosynthesis of natural products that have bioactivity, and thus potential applications as templates for new pharmaceutical products.

These advances span from the selection and breeding of elite germplasm for enhancing the production of the target phytochemical, to the recombinant production of the target phytochemical in a non-plant bioengineered host system. These are exemplified by the development of the artemisinin anti-malarial drug. Its productivity is being enhanced via traditional breeding approaches with new high-yielding hybrids to convert *A. annua* into a robust cropping system,[11, 12] and by the reconstitution of the artemisinin biosynthetic pathway in a re-engineered microbial host.[13] These long-term efforts that have been supported by the Bill and Melinda Gates Foundation (http://www.gatesfoundation.org) and are providing opportunities for entrepreneurial development of a solution to a world-health issue based upon natural product chemistry.

The literature is abundant with successes in deciphering new metabolic capabilities when this type of integrated metabolomics-transcriptomics strategy has been used in the past.[14-31] For example, the relative abundance of transcripts encoding the alkaloid biosynthetic enzymes correlates with the induction of benzylisoquinoline accumulation in *Papaver somniferum*.[14, 27] Comparison of transcriptomes and metabolomes (particularly fatty acids and lipids) of developing seeds that accumulate "unusual" fatty acids has led to the discovery of a series of FAD2-related enzymes [16-21] that are responsible for the generation of hydroxy fatty acids,[16] epoxy-fatty acids,[17, 18] conjugated fatty acids,[19, 20] and acetylenic fatty acids.[17, 21] Hence, comparing transcriptomics and metabolomics data is enabling the accurate annotation of a wide variety of genes in specialized metabolism. Due to continued technology advances, this strategy is much more readily applicable, even in the most recalcitrant systems. Database structures that enable the querying of both datasets and provide a means of generating a robust testable hypothesis concerning gene functions is important to the integration of metabolomics and transcriptomics datasets.

At present, most metabolomics databases contain metabolomics–only datasets from carefully defined samples with a common biological theme. These "themed" metabolomics databases contain intriguing and useful data,[32-40] and they make available the general framework for such databases, including the importance of data consistency and deposition of full metadata.[41] However, databases that integrate metabolomics with other global–omics data are required to provide accurate predictions of gene functions in metabolic networks. A consortium of plant biochemists, supported by the National Institutes of Health has recently completed the transcriptome and metabolome evaluation of nearly 20 medicinal plant species. The resulting datasets are accessible in publically available databases (http://medicinalplantgenomics.msu.edu/ and http://metnetdb.org/PMR/), and each offers different functionalities for integrated querying of the data in order to create testable hypotheses concerning gene-functions.[42, 43]

There is an increasing literature base that exemplifies the successful use of the Plant & Microbial Metabolomics Resource (PMR) structure and data to analyze combined transcriptomics-metabolomics datasets to identify genetic elements that support the biosynthesis of novel metabolites in medicinal plants. For example, *PMR*-enabled analyses facilitated the identification of genes associated with: a) the *Catharanthus roseus* cytochrome P450 that catalyzes a reaction in the biosynthesis of the alkaloid 19-*O*-acetylhorhammericine;[44] b) three *Digitalis purpurea* genes coding enzymes of cardenolide biosynthesis: C4 sterol methyloxidase, progesterone 5b-reductase and cardenolide synthase;[45] and c) identification of *Valeriana officinalis* genes encoding valerena-1,10-diene synthase[38] and its role in the synthesis of novel sesquiterpenes.[46] Additional analogous data from other medicinal species, such as *Echinacea purpurea* and *Hypericum perforatum* are also being used to identify genes that encode functional enzymes in the biosynthesis of unique polyketides that occur in these species (e.g., alkamides, acylphloroglucinols and hypericin).[47, 48]

3. Metabolomics based discovery and characterization of novel genes and enzymes involved in the biosynthesis of natural products

Through the correlation of gene expression and metabolite accumulation, one often obtains good hints for specific genes that are involved in the biosynthesis of related metabolites.[49] This is based on the presupposition of the mode of co-response of the biosynthetic enzyme transcripts and metabolites formed by these enzymes.[37, 50] In fact, this is a powerful strategy to identify novel genes committed in a specific biosynthetic pathway. The functions of numerous genes have been identified and characterized by metabolomics and often together with transcriptomics.[51, 52]

3.a. Flavonoids and related compounds

The genes involved in the biosynthesis and accumulation of flavonoids, some of the best known plant natural products, have been discovered with the aid of metabolomics. Naturally, the biosynthesis of specialized metabolites in *Arabidopsis thaliana* has been extensively investigated by means of integrated metabolomics and transcriptomics, because of the rich functional genomics resources available for this model plant. In *A. thaliana*, at least, 54 flavonoid molecules (35 flavonols, 11 anthocyanins and 8 proanthocyanidins) have been observed through extensive LC-MS metabolic profiling of a variety of tissues.[53] The genes responsible for the formation of scaffold structures (kaempferol, quercetin and isorhamnetin for flavonols; cyanidin for anthocyanins; and epicatechin for proanthocyanidins) have been isolated mostly through the analyses of mutants which lack pigmentation.[54, 55] However, the genes responsible for the tailoring reactions, which modify these scaffolds by glycosylation, methylation and acylation and thus subsequently responsible for the enormous chemical diversity of flavonoids, have been isolated by reverse genetic approaches where metabolomics was incorporated. Metabolic profiling of flavonoids in a series of gene-insertion mutants[56, 57] and transgenic plants overexpressing a transcription factor[58] have been used to decipher the function of genes. Generally, the predicted catalytic activities were confirmed using recombinant proteins and *in vitro* biochemical assays. This 'reverse genetics' strategy for gene identification has been most powerful when bioinformatic prediction of candidate genes obtained through a co-expression networks preceded the experimental analysis of the gene knock-out mutants. Public transcriptome datasets for *A. thaliana*, e.g. AtGenExpress,[40] are particularly useful for co-expression analyses, and usually include 'bait' genes known to be involved in a particular pathway for the discovery of other novel genes in the same co-expression framework. Several tailoring enzymes, glycosyltransferases,[59-62] acyltransferases,[63, 64] and methyltransferase[57, 65] have been characterized by this or a similar approach.

The AtMetExpress database was developed by extending the integrated analyses of gene expression and metabolic profiling data from specific Arabidopsis tissues. AtMetExpress contains LC-MS-based untargeted metabolomics data collected from a variety of *A. thaliana* tissues[66] and ecotypes.[67] Data were acquired from samples for which transcriptome and single nucleotide polymorphism data are available. From the integrated datasets, testable hypotheses could be generated. For example, the novel function of a dirigent protein was suggested to be responsible for the formation of a neolignan metabolite, and detailed experiments subsequently confirmed this function for this protein (Yonekura-Sakakibara et al., in preparation).[66]

Another database, Metabolite Profiling Database for Knock-Out Mutants in Arabidopsis (MeKO), provides information for 50 plant growth mutants, images of mutants, metabolite accumulation and interactive analysis tools.[68] Non-processed data, including chromatograms, mass spectra, and experimental metadata that follow the guidelines set by Metabolomics Standards Initiative (MSI) are freely downloadable from http://prime.psc.riken.jp/meko/. Proof-of-concept analyses suggest that the MeKO database is highly useful for gene function hypothesis generation and for improving gene annotation.

Rice (*Oryza sativa*) is a major crop critical to sustaining the human population, and as a result, rich experimental resources have been generated by past breeding programs. Metabolomics analyses have also been performed on a series of genetically defined rice inter-crossed lines to identify metabolic quantitative trait loci (QTL). QTL analyses of 87 back-crossed, inbred lines of an *indica* and *japonica* rice led to the identification and characterization of the locus on chromosome 6, which is responsible for the biosynthesis of the flavone glycoside, apigenin-6,8-di-*C*-α-L-arabinoside.[69] Using a denser genetic map including 210 recombinant inbred lines from two elite *indica* varieties, several gene candidates involved in the biosynthesis of flavone *O*- and *C*-glycosides were also characterized.[70] Using this approach, candidate genes in the narrowed QTL region could be functionally verified through additional *in vitro* biochemical assays. In particular, the genes related to the biosynthesis of flavonoid *C*-glycosides, which are characteristic metabolites in rice, were of special interest.[71]

Proanthocyanidins, also known as condensed tannins, are flavanol polymers and synthesized from the monomeric building units, flavan-3-ols (i.e. cathechin or epicathechin).[72] The molecular analyses and targeted metabolic profiling of Arabidopsis mutants exhibiting a pale yellow seed coat due to the lack of proanthocyanidins revealed the mechanism and genes involved in this biosynthetic pathway. A key enzyme, anthocyanidin reductase, catalyzes the reduction of cyanidin to epicatechin in the Arabidopsis cytosol.[73] The epicatechin is then most likely glycosylated[74] and transported into vacuole,[75] where polymerization by a laccase-like protein(s) takes place to form insoluble proanthocyanidins.[76] However, many questions regarding this process still remain to be solved.[77]

3.b. Glucosinolates

One of the first examples where metabolomics played a key role in the identification of biosynthestic genes related to plant natural products was for glucosinolate biosynthesis in Arabidopsis. Integration of *A. thaliana* transcriptome and metabolome analyses for plants subjected to sulfate-deficient stress and bioinformatic multivariate data-mining[78] enabled the pin-pointing of genes encoding a sulfotransferase[79] and a MYB transcription factor.[80] Batch-learning, self-organizing mapping (BL-SOM) analyses of concatenated matrices containing transcriptome and metabolome data was efficiently used for selecting candidate genes involved in the pathway. Co-expression network analyses with public DNA-chip microarray datasets have also been utilized as a key technology for selection of candidate genes.[81]

Genetic analysis of Arabidopsis combined with targeted analysis of glucosinolates enabled the identification of genes committed in their biosynthesis.[82] Metabolic QTL[83] and co-expression network analyses led to the identification of a novel flavin-monooxygenase responsible for *S*-oxygenation in the aliphatic glucosinolate biosynthesis.[84] Metabolic profiling of glucosinolates, camalexin and their glutathione (GSH)-conjugates in the Arabidopsis γ-glutamyl peptidase mutants resulted in the identification of two cytosolic plant γ-glutamyl peptidases involved in the processing of GSH conjugates in the glucosinolate and camalexin pathways.[85]

3.c. Terpenoids

Terpenoids are recognized as one of the most structurally-diverse classes of natural products, presumably because of the huge variation in the polymerization of isoprene units, referred as 'scaffold formation', and subsequent modification by oxygenation and glycosylation, referred to as 'tailoring reactions'. Here we discuss and provide examples of how metabolomics-based gene discovery has been used to better understand the biosynthesis of triterpene saponins. For more comprehensive discussions, readers can refer to several excellent review articles.[86-88]

β-amyrin is produced by cyclization of 2,3-oxidosqualene by specific oxidosqualene cyclases, and serves as the entry point into triterpene saponins biosynthesis. β-amyrin is then oxygenated at several carbon atoms by cytochrome P450s and these generally serve as the first 'tailoring reactions'. An elegant investigation combining expressed sequence tag analyses, gene expression analyses and metabolic profiling led to the identification two P450s, CYP88D6[89] and CYP72A154[90], which catalyze oxidation of β-amyrin to glycyrrhetinic acid in *Glycyrrhiza uralensis* (licorice). CYP88D6 catalyzes two sequential oxidations at the C-11 methylene to produce a keto-group, and CYP72A154 performs a three step oxidation at C-30 methyl to yield a carboxylic acid. (See Figure 1). A sequence homolog of CYP72A154, CYP72A63 from *Medicago truncatula*, was also able to catalyse C-30 oxidation of β-amyrin even more efficiently than CYP72A154, suggesting a function for the CYP72A subfamily proteins as triterpene-oxidizing enzymes. *M. truncatula* CYP716A12 has been suggested to be a multifunctional P450 catalyzing the oxidation at C-28 positions of β-amyrin, α-amyrin and lupeol.[91, 92] Two homologs from grape (CYP716A15 and CYP716A17) also catalyze the oxidation at C-28 of triterpenes to produce oleanolic acid, ursolic acid and betulinic acid.[92]

Glycosyltransferases catalyzing the next 'tailoring reaction' of triterpene saponins have been identified with the aid of metabolomics. From a set of co-expressed genes in *M. truncatula*, UGT73F3 was identified as an enzyme that glucosylates hederagenin at the C-28 position.[93] UGT73C10 and UGT73C11 from *Barbarea vulgaris* catalyze 3-*O*-glucosylation of the sapogenins, oleanolic acid and hederagenin.[94] Biochemical analysis of soybean mutants led to the identification of two glycosyltransferases for C-22 position of the soyasapogenol A aglycone. UGT73F2 was identified as a glucosyltransferase and UGT73F4 as a xylosyltransferase, exhibiting a close similarity as allelic genes.[95]

There are increasing lines of evidence that genes involved in some plant secondary metabolic pathways are clustered in the genome.[96, 97] The triterpenoid pathway is the best characterized. The sets of genes responsible for avenacin biosynthesis in oat (*Avena strigosa*)[98], for the synthesis of thalianol[99]- and marneral[100]-derived triterpenes in *A. thaliana*, and biosynthesis of steroidal glycoalkaloids in tomato and potato[101] have been identified as being clustered in their respective genomes. However, the recently identified *sterol side chain reductase 2* genes from tomato and potato involved in steroidal glycoalkaloid biosynthesis are located apart from these clusters.[102]

3.d. Alkaloids

The biosynthesis of morphine in opium poppy (*Papaver somniferum*) is one of the best characterized plant secondary metabolic pathways. Most genes involved in the pathway have now been identified[103, 104] by taking advantage of omics technologies including targeted-metabolite profiling. There has also been a suggestion for the presence of a gene cluster in the genome of opium poppy that may encode these metabolic functions.[105] Monoterpenoid indole alkaloids, represented by anti-neoplastic dimeric alkaloids in *Catharanthus roseus,* are formed from tryptamine and secologanin through the combination of the amino acid, tryptophan, and the monoterpenoid, iridoid, pathways. A NAD(P)H-dependent oxido-reductase-like enzyme catalyzes the formation of the iridoid scaffold.[106] A 7-deoxyleganetic acid glucosyltransferase contributes to the further steps of secolaganin biosynthesis in *C. roseus*.[107] Very recently, identification of all genes for the seco-iridoid pathway in *C. roseus* has been completed by the combined efforts of transcriptome and metabolome followed by verification by heterologous expression of candidate genes.[108, 109]

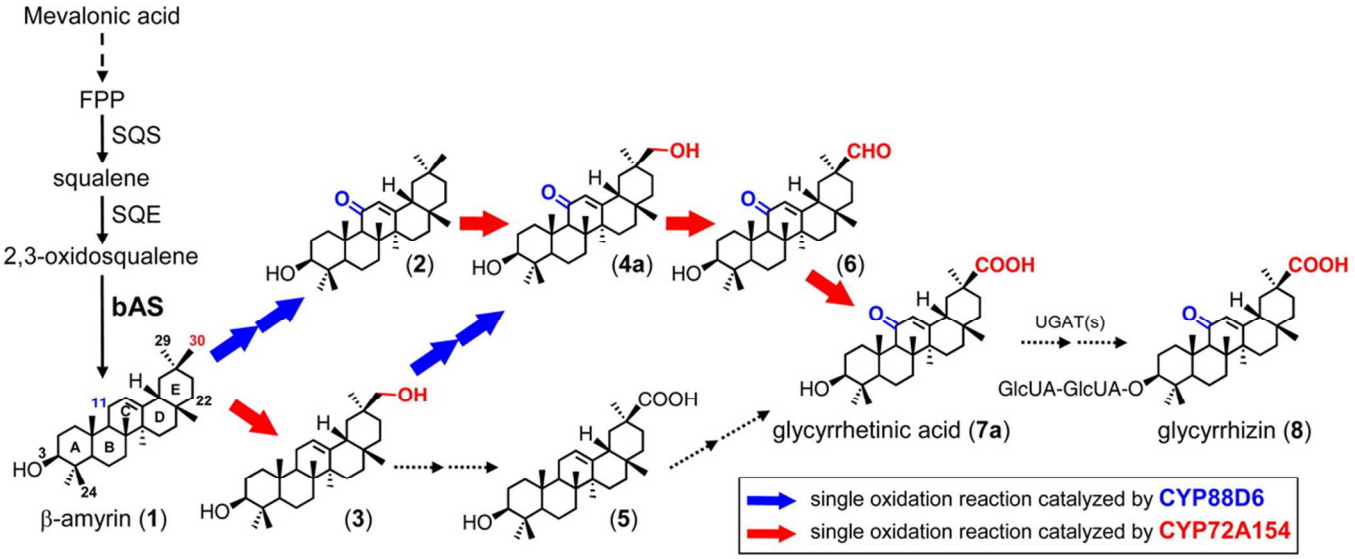

Figure 1– Proposed Pathway for Biosynthesis of Glycyrrhizin. The structures of possible biosynthetic intermediates between β-amyrin (1) and glycyrrhizin (8) are shown: (2), 11-oxo-β-amyrin; (3), 30-hydroxy-β-amyrin; (4a), 30-hydroxy-11-oxo-β-amyrin; (5), 11-deoxoglycyrrhetinic acid; (6), glycyrrhetaldehyde; and (7a), glycyrrhetinic acid. Solid black arrows indicate a dimerization reaction of two farnesyl diphosphate (FPP) molecules catalyzed by squalene synthase (SQS) originating squalene, oxidation by squalene epoxidase (SQE) to 2,3-oxidosqualene, or cyclization catalyzed by beta-amyrin synthase (bAS). A dashed arrow between mevalonic acid and farnesyl diphosphate indicates multiple enzyme reactions. The blue arrow indicates a single oxidation reaction catalyzed by the CYP88D6 enzyme (Seki et al., 2008); the red arrow indicates a single oxidation reaction catalyzed by the CYP72A154 enzyme, as described herein; the dotted arrows signify undefined oxidation and glycosylation steps. UGATs, UDP-glucuronosyl transferases. Figure from Seki H et al. Plant Cell 2011;23:4112-4123; www.plantcell.org; Copyright American Society of Plant Biologists.

An excellent example of integrated analyses of transcriptome and metabolome data for pathway elucidation of alkaloid biosynthesis is exhibited in the study of camptothecin synthesis in *Ophiorrhiza pumila* cell cultures. Differential analyses

of gene expression and metabolite accumulation between camptothecin-producing hairy roots and non-producing cell suspension cultures led to the identification of differentially accumulated metabolites and genes potentially involved in their biosynthesis.[110] Metabolomics investigation of RNA interference lines, in which the genes encoding the enzymes that catalyse the first committed reactions in the biosynthetic pathways (tryptophan decarboxylase and secologanin synthase) are suppressed, delineated the intermediary metabolite candidates in the biosynthetic pathway.[111] By taking the advantage of ultra-high resolution capacity of FT-ICR-MS, the chemical composition of the metabolomic peaks that were revealed in the LC-MS analyses were determined, and enabled the inference of the metabolite annotations following database queries.[112]

Lysine decarboxylase, which catalyzes the formation of cadaverine from lysine, is the first committed enzyme in the biosynthesis of the majority of lysine-derived alkaloids.[113] Lysine decarboxylase cDNA was identified through the comprehensive differential gene expression analyses of an alkaloid-accumulating 'bitter' variety of *Lupinus angustifolius* and an alkaloid-less 'sweet' variety.[114] The *in vivo* function of the isolated gene was confirmed by metabolomic profiling of transgenic *A. thaliana* and tobacco cells overexpressing lysine decarboxylase cDNA. Cadaverine and cadaverine-derived tobacco alkaloids were newly detected in transgenic Arabidopsis and tobacco, respectively.

4. Recent technical advances in plant metabolomics

4.a. Spatially Resolved Metabolomics

Plants as multicellular organisms integrate numerous biochemical processes that are distributed among different cell-types and among different subcellular compartments that compose a cumulative and dynamic metabolic network. The structure of the distributed metabolic networks is usually inferred by the asymmetric distribution of enzymes (proteins) or mRNAs associated with the individual components within the network. Technologies are currently available that provide high spatial and temporal resolution images of the distribution of these macromolecules within individual cells and even subcellular compartments. These imaging technologies often utilize specific macromolecular interactions between the targeted molecule and a reporter-molecule (e.g., antibodies, nucleic acid hybridizations).

Technologies that can locate the position of metabolites at such a high spatial resolution level could directly demonstrate the nature of the network and its regulation. The utility of such data is illustrated with metabolites that interact with electromagnetic waves, and are thus visibly "colored". For example, it has been known for many years that many red and blue colored anthocyanins are located in epidermal cells using cytochemical or microspectrophotometric methods,[115, 116] and this has enabled the deciphering of gene networks that program and integrate cellular differentiation, environmental cues and metabolism.[117-119] More generalized techniques have been developed that combine micro-dissection with coupling reactions that amplify the development of metabolite-dependent colored products.[120] Other methods that utilize the isolation of protoplasts[121, 122] or non-aqueous fractionation[123] of subcellular compartments have also been used to determine levels of metabolites in different compartments. However, these protocols are somewhat difficult to recapitulate and it's not always clear if these methods have "trapped", in-place the intermediates of metabolism, and thus the location of metabolites has been difficult to corroborate. More recently laser-based microdissection techniques have been applied to plant systems to harvest populations of cells that morphologically appear an identical developmental state, but most of these techniques have been applied to evaluate the transcriptome and proteome of the isolated cell populations.[124] Thus, faster and more efficient methods are needed to better analyze and understand dynamic and spatially resolved metabolism.

4.b. Mass Spectrometry Imaging

Chemical imaging is an important approach to visualizing localized metabolism. Recent technical advances have enabled metabolite imaging using mass spectrometry (MS)[125-128] and nuclear magnetic resonance (NMR).[129-132] Mass spectrometry imaging tools have also been referred to as mass microscopes and similarly, NMR imaging as NMR Microscopes.[129]

Mass spectrometry imaging (MSI) is performed through the localized ionization of metabolites, peptides, and/or proteins from specific two-dimensional[133] and more recently three-dimensional[134, 135] coordinates of a biological tissue. Thus, enabling the visualization of the spatial distribution of proteins and metabolites. A greater proportion of the current literature relates to imaging mass spectrometry of proteins, but imaging of plant small molecules and metabolites is gaining popularity[136-138].

Various ionization techniques have been used in mass spectrometry imaging. The oldest is secondary ion mass spectrometry (SIMS)[139] which was popularized for elemental and surface analyses.[140] Secondary ionization is typically achieved using a high energy monoatomic (e.g. Ga^+, Cs^+) or polyatomic (e.g. C_{60}^+) primary ion beams which are focused upon a surface, and upon impact, desorb and form secondary ions from that surface for mass analysis. The primary ions used in SIMS are typically high energy that results in substantial fragmentation of biological molecules. Thus, SIMS imaging has found the most utility in imaging of elemental/atomic species. SIMS imaging has been recently used to evaluate aluminium distribution in soybean roots,[141] elemental nutrient distribution in Phaseolus,[142] and Ar and Si in rice roots.[143] Over time, SIMS imaging of metabolites such as flavonoids in *Pisum sativum* and *A. thaliana* has been achieved using polyatomic primary ion sources[144] that distribute the energy over a broader surface and result in less damage. A primary advantage of SIMS is the ability to focus the primary ion beam either electrostatically or magnetically to achieve very high spatial resolutions. Currently nanoSIMS has been reported with a spatial resolution of 50 nanometers in plants.[145] However, SIMS is plagued by very low ionization efficiencies. Thus, it has relatively low sensitivity for biological molecules and other methods have become more popular.

Laser desorption ionization (LDI) MSI techniques are currently the most widely used techniques. Laser desorption ionization can be performed directly or assisted with a matrix. Direct LDI-MSI requires metabolites with highly absorbing chromaphores which absorb sufficient energy from the laser source to enable direct desorption and ionization. Examples of

such molecules include phenolics, flavonoids, phloroglucinols, quinones and napthoanthrones. Hölscher and colleagues recently presented a report on the use of ultraviolet-LDI to study the distribution of secondary metabolites in Arabidopsis and Hypericum species.[146] LDI using colloidal gold or silver to assist in the ionization process has also been recently reported to image epicuticular waxes from Arabidopsis leaves[147] and other tissues.[136] The Vertes group has further pioneered laser ablation electrospray ionization (LAESI),[148-154] including the development of an etched tip optical fiber that can probe single cells.[153, 154]

Matrix assisted laser desorption ionization (MALDI) is currently the most popular MSI technique due to its generally better sensitivity. MALDI-MSI was first introduced for proteins,[128] but has been rapidly adopted to small molecules[155] such as drugs[125] and plant metabolites.[138, 156] MALDI-MSI has been used in a vast number of plant applications.[123, 157-159] It has been used to image agrochemicals in soy,[160] carbohydrates in wheat,[161] lipids,[162] flavonols in apple,[163] glycoalkaloids in potato,[164] glucosinolates in Arabidopsis,[165] anthocyanins in blueberry[166] and rice,[167] and a large array of organic acids, amino acids, sugars, lipids, flavonoids and their conjugates in *M. truncatula* root nodules.[168] Although initial applications to plant systems were focused on metabolites that occur on the exterior surfaces of plants (i.e., epicuticular lipids)[169-171], histological preparative methods are also being evaluated to image "interior" metabolites of plant tissues.[136, 172]

Desorption electrospray ionization (DESI) is another ionization technique that is being developed[173-176] and has found diverse applications in metabolite imaging.[157, 177-184] DESI imaging is achieved by directing an electrospray aerosol onto an ambient surface. Secondary ions originating from the surface are then generated and used for mass spectral imaging. Alternatively, plant metabolites can be transferred or imprinted onto a separate porous polytetrafluoroethylene (PTFE) surface and similarly imaged from the PTFE surface to enhance the signal intensity.[157] DESI-MSI has been used to image diterpene glycosides in Stevia leaves[185] and alkaloids in various tissues of poisonous hemlock (*Conium maculatum*), jimsonweed (*Datura stramonium*) and nightshade (*Atropa belladonna*).[186] Laskin and colleagues dramatically improved the spatial resolution of DESI imaging, using a nanoDESI technique and achieved spatial resolution of about 12 µm.[187]

The spatial resolution of early MALDI-MSI was approximately 150 x 150 µm, which is multi-cellular for most organisms. The spatial resolution of MALDI-MSI is predominantly dictated by the laser optics and spot size. However, improvements in laser optics have resulted in improved spatial resolutions, and many commercial systems are now available with a resolution of approximately 10 x 10 µm. Thus, there has been an almost 10-fold increase in the spatial resolution to position imaged metabolites relative to known cellular structures, from approximately 100-µm resolution in 2005,[171] to approximately 10-µm resolution in 2010.[169, 188] Considering that plant cells are approximately 30-50 µm in size, this level of spatial resolution is enabling the localization of metabolites to the level of individual cellular structures and the spatial resolution continues to increase. An ultra-high spatial resolution version of matrix-assisted laser desorption ionization (MALDI) method, called scanning microprobe MALDI (SMALDI) has been developed, and it can determine the position of analytes to a resolution of few microns[189, 190]. Other custom MSI instruments with resolutions of 1 µm are also being built in research labs, but commercial systems with comparable high spatial resolutions are not yet readily available.

As the spatial resolution increases due to decreased laser spot size, the number of 'pixels' needed to image the same area increases and so does the imaging time. The increased imaging times have been mediated by additional improvements/increases in laser frequencies with 1 kHz lasers now being common. However, continual improvemetns in increased spatial resolution and imaging speeds are still desired to enable greater resolution of smaller cells and sub-cellular metabolites in shorter times. Unfortunately, we may be nearing the practical spatial limits of current MSI technology due to the reduced quantities of metabolites in smaller volumes (i.e., sensitivity), and the Rayliegh limits of the optical UV lasers, which are on the order of several hundred nanometers. Current MSI also faces the challenges of competitive ionization and ion suppression making it more difficult.

4.c. NMR Imaging

Nuclear magnetic resonance imaging, or often referred to as simply magnetic resonance imaging (MRI), is another chemically specific, spatial imaging technology which measures the resonance intensity of paramagnetic nuclei found in many biological metabolites to generate images.[191] See Ishida *et al.*[129] and Kockenberger *et al.*[130] for earlier reviews of MRI. One of the major benefits of MRI is that it is non-destructive and can image intact, living plant tissues. MRI is achieved by measuring the resonance intensity while varying a magnetic gradient along the x and y axes at spatially fixed values.[129-131] During this process, the resonance signal is recorded for only a small spatially distinct part of the sample and sensitivity can be a significant challenge. As the desired spatial resolution increases, the signal to noise ratio is further decreased. Thus, many MRI experiments utilize the ^{1}H resonance signal from water due to its relatively high abundance resulting in high sensitivity. As a result, MRI has been extensively used to study water content and movement/transport throughout plants.[192-196] The abundant ^{1}H resonance signal from water is also preferred and necessary to obtain the highest resolution images. Theoretical resolutions of up to 1 µm have been predicted for MRI with realistic three dimensional resolutions of 3.7 x 3.3 x 3.3 μm^3 achieved.[197] It is likewise possible to image ^{1}H signals originating from metabolites such as sugars and amino acids,[198, 199] as well as to image other paramagnetic nuclei such ^{13}C, ^{15}N, ^{17}O, ^{19}F, ^{23}Na, ^{31}P, and ^{39}K.[200, 201] Some of the most impressive recent MRI of plants has been focused on imaging lipid distribution in seeds.[202, 203]

MRI is a well-developed technology that continues to improve with the improved sensitivities enabled through higher-field magnets and cryogenically cooled probes. However, its current resolution is on the order of 1-10 µm in the x, y and z planes and varies dependent upon the relative sensitivity of the selected nuclear resonance signal. Similar to mass spectrometry imaging, the primary limitation to MRI resolution is the low concentration of metabolites localized within smaller spatial regions. Both MSI and MRI can achieve approximate cellular resolution for many plants, but their spatial resolutions ultimately fall short of most optical imaging technologies. However, both MSI and MRI provide chemically

specific spatial details that simply cannot be achieved with optical techniques. Thus, both MSI and MRI are contributing to our understanding of the differential and dynamic metabolism that is occurring throughout plants at an ever growing rate, and are destined to be key metabolomics technologies in the future. MSI has the current advantage that it can differentially profile and image a larger number of m/z values or potentially different metabolites simultaneously relative to MRI, but MRI has the predominant advantage that it is non-destructive and can image intact, living organisms.

4.d. LC-MS-SPE-NMR

The large-scale profiling of plant metabolites (i.e. metabolomics) is advancing our fundamental understanding of plant biochemistry, yielding discovery of novel metabolites and gene functions, and providing an advanced mechanistic understanding of plant responses to biotic, abiotic, and environmental stimuli. See Section 3 for details. However, the current depth-of-coverage is still limiting and chemical annotation is the number one challenge for the Metabolomics Community.[204-206] Thus, there is a critical need to identify more metabolites both systematically (i.e. logical progression towards identification of all metabolites) and in a biologically directed manner (i.e. those observed to be differentially accumulated during comparative experiments) so as to increase our depth-of-coverage and to enhance the biological context and content of metabolomics experiments.

The Metabolomics Community has come to a general consensus that a minimum of two orthogonal analytical data are necessary for confident identification relative to an authentic standard.[207] In the absence of authentic standards, it is necessary to return to the traditional standards of the organic chemistry journals that include NMR and accurate mass measurements for elemental analyses. Historically, mass spectrometry and NMR workflows have been segregated, and the identification of metabolites by NMR has been a lengthy process. However, MS and NMR methods are now being brought together in a more integrated manner.

One approach is to couple HPLC directly to NMR by transferring the HPLC eluent to a NMR flow-probe that can be operated in continuous-flow or stopped-flow manner.[208] However, the practical implementation and utility of these approaches is challenging due to low sensitivities, reduced time for NMR spectral acquisition, and high costs of deuterated HPLC solvents. These limitations can be circumvented through the automated purification and concentration of one or many HPLC separated metabolites via post-column, solid-phase extraction (SPE). HPLC is often coupled to parallel or in-line detectors such as UV and MS resulting in a complex instrumental ensemble, *i.e.* HPLC-UV-MS-SPE (Figure 2). Purified samples collected from repetitive HPLC-UV-MS-SPE separations can then be eluted with deuterated solvents into traditional NMR tubes or into a flow-probe for NMR analyses.[209-211]

The authors of this article are convinced that such integrated MS and NMR techniques are necessary for 'higher' through-put metabolite identifications and large-scale plant metabolomics. Thus, the authors are advocating this concept by integrating ultra-high pressure liquid chromatography (UHPLC, 18,000 psi) with high resolution quadrupole time-of-flight mass spectrometry (QTofMs, resolution ~ 40,000) and SPE for the automated and higher through-put purification and concentration of chemically undefined plant metabolites. Specific metabolites targeted for annotation are repetitively collected from multiple UHPLC-UV-MS separations onto the same SPE cartridge which allows for the recovery of larger quantities necessary for NMR analyses. This system enables high resolution separations of metabolites that can then be further isolated using a combination of mass, retention time, and UV signals to initiate SPE purification and concentration. Purified and concentrated target metabolites isolated on SPE cartridges are dried to remove protonated solvents and eluted with deuterated solvents for further NMR analyses. Structural identifications are then made from the combination of UHPLC retention, UV, MS, MS/MS, and NMR data. Targeted recovery amounts of 1-5 μg are typically sufficient for 1D and 2D NMR using a Bruker 600 MHz NMR with a 1.7 mm TCI cryoprobe.

Journal Name RSCPublishing

ARTICLE

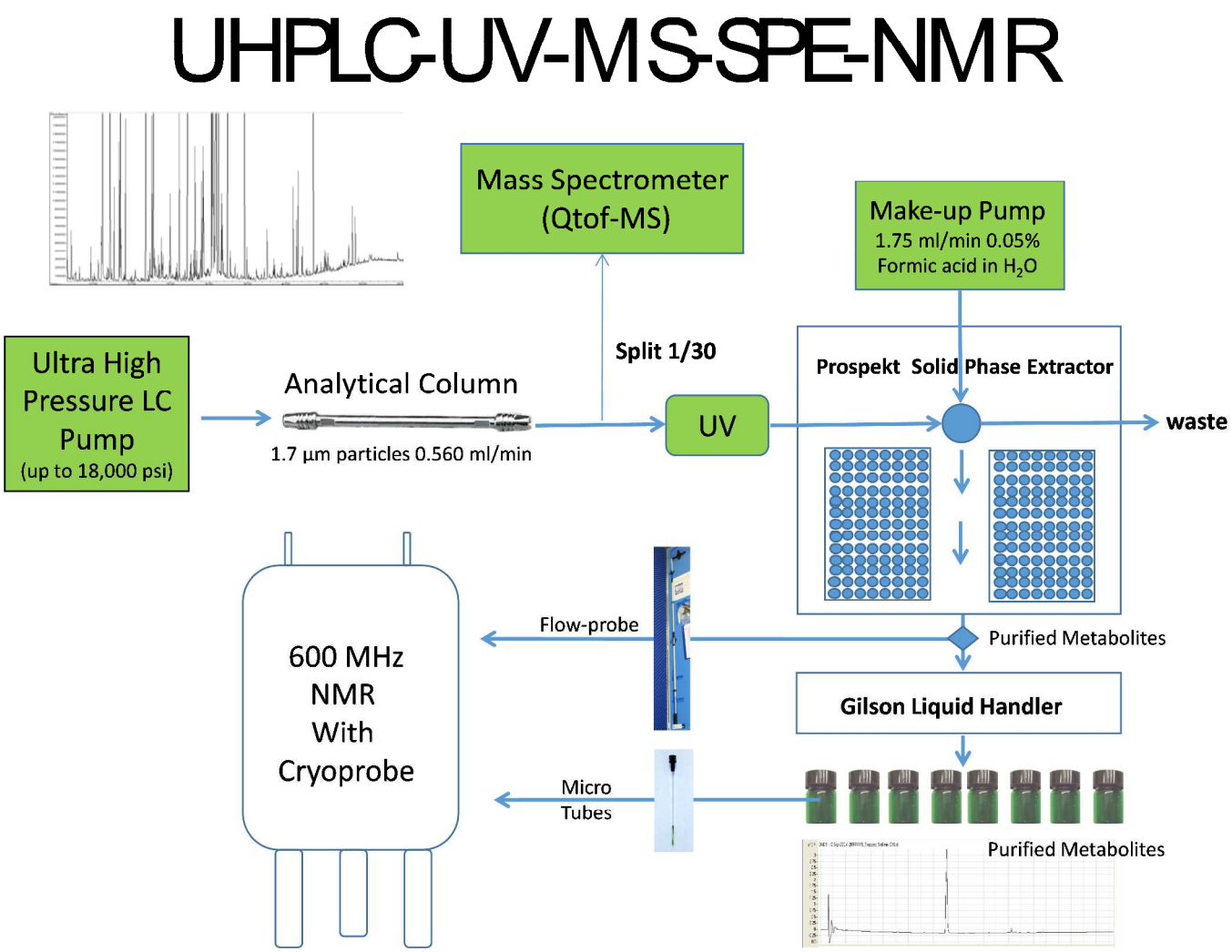

Figure 2 – Schematic of an integrated UHPLC-UV-MS-SPE-NMR system. Comparative metabolomics is performed using large-scale metabolite profiling by UHPLC-UV-MS. Differentially accumulated metabolites or otherwise targeted metabolites are then purified and concentrated by solid-phase extraction (SPE) from multiple UHPLC-UV-MS separations. The UHPLC-UV-MS peaks of interest can be targeted for SPE purification based upon retention time, m/z, and UV absorbance. One or many targeted metabolite/peaks can be repetitively isolated on different SPE columns (up to 2 x 96 using a Bruker/Spark Holland Prospekt 2). SPE purified and concentrated metabolites are then dried to remove protonated solvents and eluted using deuterated solvents into traditional tubes or into a flow-probe for NMR analyses. Metabolites are then ultimately identified based upon retention, UV absorbance, accurate mass, tandem mass, 1D and/or 2D $^{1}H/^{13}C$ NMR spectra.

4.e. Tandem Mass Spectral Database and Computational MS

As noted above, the major challenge in plant metabolomics is metabolite annotation. Typically thousands of metabolites (or mass features) can be detected in an untargeted metabolomics experiment. Identification of those metabolites is very challenging because (1) many natural products have a substantial number of isomers which cannot be differentiated solely based upon their m/z values, and (2) for each single mass feature, multiple empirical formulas can often be generated within a small m/z variance window, *i.e.*, < 5 ppm. Tandem mass spectrometry is therefore often employed to provide additional structural information that can be used to annotate metabolites by matching tandem spectra of experimental samples to those of authentic compounds within a database. Over the past years, a number of online public tandem mass spectral databases have been developed. These include MassBank,[212] HMDB,[213] Metlin,[214, 215] Golm Metabolome Database (GMD),[216] the Platform for RIKEN Metabolomics

(PRIMe),[217, 218] MeltDB,[219] and Madison Metabolomics Consortium Database (MMCD).[220] Many of these databases focus upon specific organisms. For example, the PRIMe database focuses primarily on plant metabolomes, the HMDB is dedicated more towards the human metabolome, and Metlin is directed more towards pharmaceutical research and biomarker discovery. Among these public databases, MassBank is unique in that it also serves as a public repository for the metabolomics and the natural product research communities. It currently has more than 40,000 spectra contributed by 28 different institutes worldwide. It is also one of the first public mass spectral data repositories that allow users to share their spectral data. This is very important as many "unknowns" encountered in untargeted metabolomics are often "known unknowns". Identification of these "known unknowns" can be difficult as their structural information is often scattered throughout the literature and various databases, some of which may be difficult to assess. Providing a public platform for researchers to deposit and share their mass spectral data can significantly improve metabolite annotation within the plant metabolomics community. In addition to public databases, commercial databases are also available such as the Agilent Fiehn GC-MS Metabolomics Library, Agilent Metlin Personal Metabolite Database and Wiley Registry/NIST Mass Spectral Library that now contains both electron ionization (EI) and tandem MS/MS spectra. Unlike EI spectra, electrospray ionization MS/MS spectra are historically less reproducible. Different instrument types (or even the same instrument type but from different vendors), different ion optics and different collision energies can produce substantial variations within the MS/MS spectra.[221] This often leads to unsuccessful identifications and even misidentifications. In addition, the overall metabolome coverage of the available MS/MS databases is still very limited due to the lack of authentic standards and the complexity of the metabolomes. Thus, the identification of metabolites not included in the databases is challenging and typically requires additional computational MS (or MS informatics) or empirical methods, ie NMR.

Computational annotation tools based upon MS data have become an important part of the metabolomics workflow owing to their rapid developments over the past decade.[222] *In silico* generated MS/MS databases are expected to significantly improve the success of metabolite annotation when experimental MS/MS training data of authentic standards become available. Several public or commercial software packages and websites are now available, including MetFrag and Metfusion,[223, 224] Fragment Identificator (FiD),[225] lipidblast,[226] SIRIUS,[227] MS interpreter (NIST),[228] MS fragmenter (ACD lab) and Mass Frontier (HighChem). *In silico* fragmentation is predicted based upon fragmentation rules generalized from literature data[229, 230] or combinatorial approaches that use bond energies to predict fragmentation chemistries.[223-225] While rule-based approaches can be highly specific, especially for the same class of metabolites, they typically cannot correctly predict fragments for metabolites for which fragmentation rules are absent. Combinatorial methods require extensive computation when the candidate metabolite has a large number of chemical bonds and functional groups, and they lack the specificity of the rule-based approaches. Some software, such as Mass Frontier, contain manually curated fragmentation patterns from several thousand publications. *In silico* fragmentation prediction has been successfully used with compounds consisting of similar repeating units such as peptides,[231] lipids,[226, 232] and glucans.[233, 234] However, its use on arbitrary small molecules is still very challenging due to the structural diversity of small molecules and their complex fragmentation patterns. For example, we observed that hydroxylation position renders fragmentation of structurally similar flavonoids completely different. Thus an algorithm that combines rule-based approaches, combinatorial methods, published fragmentation mechanisms and machine learning are likely to be more successful in predicting fragmentation of small molecules.

4.f. Large-scale targeted metabolomics via MRM

Large-scale targeted metabolomics aims to detect and quantify dozens to hundreds of known compounds in a complex sample mixture. It can be used to assess metabolic changes resulting from genetic manipulation and/or environmental perturbation by selectively monitoring a subset of metabolites associated with certain specific pathways. Ultrahigh performance liquid chromatography coupled to a triple quadrupole MS (UPLC-QqQ-MS) is ideal for targeted metabolomics due to its good sensitivity, reproducibility, robust quantification and broad dynamic range. It is typically operated in MRM (Multiple Reaction Monitoring) mode in which collision energies and other parameters for each individual target compounds have been pre-optimized with authentic standards to enhance sensitivity and selectivity.[235] During a MRM experiment, a metabolite precursor is first resolved and isolated by the first quadrupole (Q1) and fragmented in the second quadrupole (Q2) which functions as a collision cell. The third quadrupole (Q3) serves as a final mass filter to monitor specific product ions. The identity of metabolites can be ensured when combined with known chromatographic retention times because individual metabolites have specific precursor/product ions, which are also known as a transition pair. Quantification is performed using the more specific and abundant transition pair for each targeted metabolite.

Large-scale MRM has become a workhorse for targeted metabolomics due to its high sensitivity and selectivity. The utility of MRM in targeted plant metabolomics has been well demonstrated.[236-238] Typically, two MRM transitions, *i.e.,* a quantitative transition and a qualitative transition, are monitored to increase the confidence in compound annotation. However, due to the complexity of the metabolomes, some isomeric metabolites might not be adequately separated by LC and they might produce the same product ions that are used for monitoring the target compounds.[235] This can lead to false positive results.[239, 240] Some solutions, ranging from the use of probability-based computational approaches to a different instrument setup, have been proposed to minimize reporting of false positive results and confirmation of identities.[241, 242] Triggered MRM or MRM-EPI (MRM-triggered enhanced product ion scan) that initiate the acquisition of a product ion spectrum when the signal of the MRM transition exceeds a pre-set threshold can be very useful in confirming the identity of the metabolites.[243] However, if a product ion spectrum is acquired for every target metabolite, the total number of transitions can increase substantially. This can lead to poor quantification accuracy and low sensitivity due to compressed acquisition/dwell times for the individual MS/MS transitions. Another challenge in MRM targeted metabolomics is that the product ion spectra can be dominated by one major fragment for some metabolites. Other fragments have very low signal

intensities and thus are less selective and sensitive if used as MRM transitions. In addition, for some metabolites, neutral losses of H_2O, CO_2, methyl or glycosyl groups are the major fragment ions. Ions resulting from these neutral losses are less selective and commonly give rise to matrix interferences.

4.g. X-ray using porous complexes

The use of X-ray crystallography in structural determinations has a long history, and the first crystallographic structural elucidation of a small organic molecule was performed in 1923.[244] Early use of X-ray crystallography focused on inorganic compounds and minerals. However, it's utility in the determination of organic and biological relevant molecules eventually evolved, and the Noble Prize in Chemistry was awarded to Dorothy C. Hodgkin in 1964 for solving the structures of important biological molecules such as cholesterol, penicillin and vitamin B12.[245] However, over the past 20 years or so, crystallography has fallen from favour as a structural determination tool for small molecules. This has been primarily due to the lengthy time needed to generate relatively large quantities of the analyte required for the often numerous trial-and-error attempts to produce quality crystals relative to other technologies such as MS and NMR. In addition, not all molecules can form the crystalline structures necessary for X-ray analysis.

Fortunately, transformative new technology is now making X-ray structural determination of µg to ng quantities of biomolecules a reality.[246, 247] Inokuma and colleagues have developed a method that removes the bottleneck needed to produce relatively large single crystals for single-crystal X-ray diffraction (SCD) analyses and allows for the structural determination of non-crystalline materials. Their method instead dissolves targeted analytes in a solvent that is then slowly infused into a porous, crystalline host complex. As the analyte infuses into the host structure it absorbs onto the pore surfaces via noncovalent chemical interactions. This produces an ordered structure of the analyte within the ordered host crystal as the solvent is slowly evaporated. The structure of the analyte is then solved within the structure of the crystalline host matrix using traditional instrumentation. The authors report that their standard protocol utilizes approximately 5 µg of analyte, but further demonstrate their approach by solving the crystal structure from as little as 80 ng of guaiazulene using standard laboratory instrumentation; *i.e.,* Bruker APEX-II CCD with a Mo radiation source or Rigaku VariMax with a Cu radiation source. The authors further propose that structures could be determined from as little as 10ng or less using a synchrotron radiation source.

Inokuma and colleagues additionally demonstrated that their method was compatible with HPLC purified natural products from orange peel (*Citrus unshiu*).[246] In this approach, host crystal 'sponges' were added directly to HPLC fractions collected in vials followed by subsequent X-ray crystallography. Using their HPLC-SCD approach, they determined the structures of three polymethoxyflavones and concluded that '*LC–SCD analysis will be a powerful tool for the rapid characterization of multiple components with much higher structural reliability than liquid chromatography mass spectrometry and LC–NMR techniques*'.

Overall, the sensitivity of the reported HPLC-SCD technique rivals that of mass spectrometry and surpasses that of NMR. Thus, it is very conceivable that HPLC-SCD could be a powerful and realistic alternative for higher through-put metabolite identifications. Integration of SCD with HPLC-MS-SPE-SCD is also easily visualized.

5. Future prospects for crop breeding and herbal medicine

A large proportion of the success of plant metabolomics has been obtained using model species. However, there are great needs, opportunities and challenges associated with expanding the utility of metabolomics in crop species. A major challenge is the complex genetics generally associated with crops. However, metabolomics has successfully used for enhanced breeding of important crops.[248] As a basis for future breeding programs, the identification of genomic regions and genes associated with metabolic quantitative loci (mQTL) or production of specialized metabolites has been performed with major crops taking advantage of available genetic resources. Extensive mQTL[69, 249], phenotypic/genetic[250, 251] analyses and genome wide association study (GWAS)[252] of rice populations have been reported, noting that rice is a primary crop feeding the majority of the world's population. In these studies, the mQTL regions[69, 249] and related single-nucleotide polymorphism (SNP) markers[252] for a variety of flavonoid molecules have been identified. These metabolomic investigations were based on the large-scale identification of rice metabolites using a solid strategy of structure determination by natural product chemistry.[253] Besides the direct relation of genomic regions or genes with metabolites, investigation of metabolic signatures representing the responses to abiotic stresses is also an important issue being addressed through metabolomics. For example, metabolomics studies revealed characteristic metabolic changes associated with the stresses of draught, UV and nitrogen deficiency in wheat[254, 255] and maize.[256, 257]

The application to herbal medicines and crude drugs is also expected to benefit from metabolomics analyses.[258, 259] Metabolomics can depict not only large numbers of chemical components found in mixtures of herbal medicines,[260] but can also correlate those chemical components from plants with the chemical markers of patients who intake these herbal medicines. If one can systematize all correlations of chemical components both from herbal medicines and body fluids of patients, such as blood and urine with diagnostic indices, new prescriptions of herbal medicine mixtures could be developed to maximize the therapeutic effects.[261]

6. Conclusions

Metabolomics has significantly advanced our understanding of plant specialized metabolism and natural product biosynthesis at the molecular and biochemical levels with numerous examples provided here within. Metabolomics is also enabling the better understanding of medicinal plants and the identification of important metabolic QTLs for enhanced breeding. Although metabolomics has proven its value, it still faces substantial challenges including large-scale metabolite identifications. As improved technologies continue their

progressive march forward, the field of metabolomics can only get better. A large number of advancing technologies were reviewed within this article and provide a perspective on the exciting and growing potential of metabolomics in the future!

7. Acknowledgements

All the authors would like to acknowledge support from the Metabolomics for a Low Carbon Society program funded jointly by the US National Science Foundation IOS (Award #1139489) and the Japanese Science and Technology Agency (SICORP). The authors would also like to thank the National Science Foundation IOS Award#1340058 for funding a related Research Coordination Grant: Integrating and Coordinating a National and International Plant, Algae, and Microbial Metabolomics Research Coordination Network that is enabling greater interactions between the international scientists funded through the joint NSF-JST Metabolomics for a Low Carbon Society Program. LWS would like to acknowledge both the National Science Foundation, Major Research Instrumentation, DBI Award #1126719 and Bruker Daltonics/BioSpin; especially Aiko Barsch and Ulrich Braumann, for UHPLC-MS-SPE-NMR instrumentation support. BJN acknowledges the on-going support by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences through the Ames Laboratory. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358.

Notes and References

[a] The Samuel Roberts Noble Foundation, Plant Biology Division, 2510 Sam Noble Parkway, Ardmore, OK USA.

[b] Dept. of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011.

[c] Ames Laboratory, U.S. Department of Energy, Ames, IA 50011

[d] RIKEN Center for Sustainable Resource Science, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

[e] Graduate School of Pharmaceutical Sciences, Chiba University, Inohana, Chuo-ku, Chiba 260-8675, Japan.

1. I. F. Stone, *The trial of Socrates*, Little, Brown, Boston, 1988.
2. B. Robson, O. K. Baek and S. Ekins, *The Engines of Hippocrates: From the Dawn of Medicine to Medical and Pharmaceutical Informatics*, Wiley, 2009.
3. E. J. DaSilva, E. Baydoun and A. Badran, *Biotechnology and the developing world*, 2002.
4. W. Sneader, *BMJ*, 2000, **321**, 1591-1594.
5. D. Greenwood, *Journal of Antimicrobial Chemotherapy*, 1992, **30**, 417-427.
6. M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon and A. T. McPhail, *Journal of the American Chemical Society*, 1971, **93**, 2325-2327.
7. J. Goodman and V. Walsh, *The story of taxol : nature and politics in the pursuit of an anti-cancer drug*, Cambridge University Press, Cambridge ; New York, 2001.
8. L. H. Miller and X. Su, *Cell*, 2011, **146**, 855-858.
9. S. J. Cutler and H. G. Cutler, *Biologically active natural products : pharmaceuticals*, CRC Press, Boca Raton, FL, 2000.
10. J. R. Hanson, *Natural products : the secondary metabolites*, Royal Society of Chemistry, Cambridge, 2003.
11. I. A. Graham, K. Besser, S. Blumer, C. A. Branigan, T. Czechowski, L. Elias, I. Guterman, D. Harvey, P. G. Isaac, A. M. Khan, T. R. Larson, Y. Li, T. Pawson, T. Penfield, A. M. Rae, D. A. Rathbone, S. Reid, J. Ross, M. F. Smallwood, V. Segura, T. Townsend, D. Vyas, T. Winzer and D. Bowles, *Science*, 2010, **327**, 328-331.
12. T. Townsend, V. Segura, G. Chigeza, T. Penfield, A. Rae, D. Harvey, D. Bowles and I. A. Graham, *PloS one*, 2013, **8**, e61989.
13. V. Hale, J. D. Keasling, N. Renninger and T. T. Diagana, *Am J Trop Med Hyg*, 2007, **77**, 198-202.
14. A. Ounaroon, G. Decker, J. Schmidt, F. Lottspeich and T. M. Kutchan, *Plant J.*, 2003, **36**, 808-819.
15. J. M. Gennity and P. K. Stumpf, *Arch Biochem Biophys*, 1985, **239**, 444-454.
16. M. Bafor, M. A. Smith, L. Jonsson, K. Stobart and S. Stymne, *Biochem. J.*, 1991, **280**, 507-514.
17. M. Lee, M. Lenman, A. Banas, M. Bafor, S. Singh, M. Schweizer, R. Nilsson, C. Liljenberg, A. Dahlqvist, P. O. Gummeson, S. Sjodahl, A. Green and S. Stymne, *Science*, 1998, **280**, 915-918.
18. M. Bafor, M. A. Smith, L. Jonsson, K. Stobart and S. Stymne, *Biochem. Biophys.*, 1993, **303**, 145-151.
19. E. B. Cahoon, K. G. Ripp, S. E. Hall and A. J. Kinney, *J. Biol. Chem.*, 2001, **276**, 2637-2643.
20. J. M. Dyer, D. C. Chapital, J. C. Kuan, R. T. Mullen, C. Turner, T. A. McKeon and A. B. Pepperman, *Plant Physiol.*, 2002, **130**, 2027-2038.
21. P. Sperling, M. Lee, T. Girke, U. Zahringer, S. Stymne and E. Heinz, *Eur. J. Biochem.*, 2000, **267**, 3801-3811.
22. T. Cakir, K. R. Patil, Z. Onsan, K. O. Ulgen, B. Kirdar and J. Nielsen, *Molecular systems biology*, 2006, **2**, 50.
23. D. B. Murray, M. Beckmann and H. Kitano, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 2241-2246.
24. M. T. Kresnowati, W. A. van Winden, M. J. Almering, A. ten Pierick, C. Ras, T. A. Knijnenburg, P. Daran-Lapujade, J. T. Pronk, J. J. Heijnen and J. M. Daran, *Molecular systems biology*, 2006, **2**, 49.
25. M. A. Smith, L. Jonsson, S. Stymne and K. Stobart, *Biochem. J.*, 1992, **287**, 141-144.
26. N. Amiour, S. Imbaud, G. Clement, N. Agier, M. Zivy, B. Valot, T. Balliau, P. Armengaud, I. Quillere, R. Canas, T. Tercet-Laforgue and B. Hirel, *J. Exp. Bot.*, 2012, **63**, 5017-5033.
27. T. T. Dang and P. J. Facchini, *Plant Physiol.*, 2012, **159**, 618-631.
28. A. Jauhiainen, O. Nerman, G. Michailidis and R. Jornsten, *Biostatistics*, 2012, **13**, 748-761.
29. S. Osorio, R. Alba, Z. Nikoloski, A. Kochevenko, A. R. Fernie and J. J. Giovannoni, *Plant Physiol.*, 2012, **159**, 1713-1729.
30. S. P. Singh, X. R. Zhou, Q. Liu, S. Stymne and A. G. Green, *Curr. Opin. Plant Biol.*, 2005, **8**, 197-203.
31. Y. Gibon, B. Usadel, O. E. Blaesing, B. Kamlage, M. Hoehne, R. Trethewey and M. Stitt, *Genome biology*, 2006, **7**, R76.
32. P. Bais, S. M. Moon, K. He, R. Leitao, K. Dreher, T. Walk, Y. Sucaet, L. Barkan, G. Wohlgemuth, M. R. Roth, E. S. Wurtele, P. Dixon, O. Fiehn, B. M. Lange, V. Shulaev, L. W. Sumner, R. Welti, B. J. Nikolau, S. Y. Rhee and J. A. Dickerson, *Plant Physiol.*, 2010, **152**, 1807-1816.
33. S. M. Quanbeck, L. Brachova, A. A. Campbell, X. Guan, A. Perera, K. He, S. Y. Rhee, P. Bais, J. A. Dickerson, P. Dixon, G. Wohlgemuth, O. Fiehn, L. Barkan, I. Lange, B. M. Lange, I. Lee, D. Cortes, C. Salazar, J. Shuman, V. Shulaev, D. V. Huhman, L. W. Sumner, M. R. Roth, R. Welti, H. Ilarslan, E. S. Wurtele and B. J. Nikolau, *Front. Plant Sci.*, 2012, **3**, 15.
34. D. W. J. James, E. Lim, J. Keller, I. Plooy, E. Ralston and H. K. Dooner, *Plant Cell*, 1995, **7**, 309-319.
35. C. S. Oh, D. A. Toke, S. Mandala and C. E. Martin, *J. Biol. Chem.*, 1997, **272**, 17376-17384.
36. M. N. Ngaki, G. V. Louie, R. N. Philippe, G. Manning, F. Pojer, M. E. Bowman, L. Li, E. Larsen, E. S. Wurtele and J. P. Noel, *Nature*, 2012, **485**, 530-533.
37. H. Goda, E. Sasaki, K. Akiyama, A. Maruyama-Nakashita, K. Nakabayashi, W. Li, M. Ogawa, Y. Yamauchi, J. Preston, K. Aoki, T. Kiba, S. Takatsuto, S. Fujioka, T. Asami, T. Nakano, H.

Kato, T. Mizuno, H. Sakakibara, S. Yamaguchi, E. Nambara, Y. Kamiya, H. Takahashi, M. Y. Hirai, T. Sakurai, K. Shinozaki, K. Saito, S. Yoshida and Y. Shimada, *Plant J.*, 2008, **55**, 526-542.
38. E. S. Wurtele, J. Chappell, A. D. Jones, M. D. Celiz, N. Ransom, M. Hur, L. Rizshsky, P. Dixon, J. Liu, M. P. Widrlechner and B. J. Nikolau, *Metabolites*, 2012, In Press.
39. M. C. Crispin and E. S. Wurtele, in *Biotechnology for Medicinal Plants*, eds. S. Chandra, H. Lata and A. Varma, Springer Berlin Heidelberg, , 2013, pp. 395-411.
40. H. Goda, E. Sasaki, K. Akiyama, A. Maruyama-Nakashita, K. Nakabayashi, W. Li, M. Ogawa, Y. Yamauchi, J. Preston, K. Aoki, T. Kiba, S. Takatsuto, S. Fujioka, T. Asami, T. Nakano, H. Kato, T. Mizuno, H. Sakakibara, S. Yamaguchi, E. Nambara, Y. Kamiya, H. Takahashi, M. Y. Hirai, T. Sakurai, K. Shinozaki, K. Saito, S. Yoshida and Y. Shimada, *The Plant journal : for cell and molecular biology*, 2008, **55**, 526-542.
41. H. Jenkins, N. Hardy, M. Beckmann, J. Draper, A. R. Smith, J. Taylor, O. Fiehn, R. Goodacre, R. J. Bino, R. Hall, J. Kopka, G. A. Lane, B. M. Lange, J. R. Liu, P. Mendes, B. J. Nikolau, S. G. Oliver, N. W. Paton, S. Rhee, U. Roessner-Tunali, K. Saito, J. Smedsgaard, L. W. Sumner, T. Wang, S. Walsh, E. S. Wurtele and D. B. Kell, *Nat. Biotechnol.*, 2004, **22**, 1601-1606.
42. E. Wurtele, J. Chappell, A. Jones, M. Celiz, N. Ransom, M. Hur, L. Rizshsky, M. Crispin, P. Dixon, J. Liu, M. P.Widrlechner and B. Nikolau, *Metabolites*, 2012, **2**, 1031-1059.
43. M. Hur, A. A. Campbell, M. Almeida-de-Macedo, L. Li, N. Ransom, A. Jose, M. Crispin, B. J. Nikolau and E. S. Wurtele, *Natural product reports*, 2013, **30**, 565-583.
44. L. A. Giddings, D. K. Liscombe, J. P. Hamilton, K. L. Childs, D. DellaPenna, C. R. Buell and S. E. O'Connor, *J. Biol. Chem.*, 2011, **286**, 16751-16757.
45. Y. S. Yeo, S. E. Nybo, A. G. Chittiboyina, A. D. Weerasooriya, Y. H. Wang, E. Góngora‐Castillo, B. Vaillancourt, C. R. Buell, D. DellaPenna, M. D. Celiz, A. D. Jones, E. S. Wurtele, N. Ransom, N. Dudareva, K. A. Shabaan, N. Tibrewal, S. Chandra, T. Smillie, I. A. Khan, R. M. Coates, D. S. Watt and J. Chappell, *J. Biol. Chem.*, 2012, In Review.
46. S. Takahashi, Y. Yeo, B. T. Greenhagen, T. McMullin, L. Song, J. Maurina-Brunker, R. Rosson, J. P. Noel and J. Chappell, *Biotechnology and bioengineering*, 2007, **97**, 170-181.
47. M. Crispin and E. Wurtele, in *Biotechnology for Medicinal Plants*, eds. S. Chandra, H. Lata and A. Varma, Springer Berlin Heidelberg, 2013, pp. 395-411.
48. M. C. Crispin, M. Hur, T. Park, Y. H. Kim and E. S. Wurtele, *Physiologia plantarum*, 2013, **148**, 354-370.
49. K. Saito, *Current opinion in plant biology*, 2013, **16**, 373-380.
50. A. R. Fernie, *Phytochemistry*, 2007, **68**, 2861-2880.
51. K. Saito and F. Matsuda, *Annu Rev Plant Biol*, 2010, **61**, 463-489.
52. K. Yonekura-Sakakibara and K. Saito, *Natural product reports*, 2009, **26**, 1466-1487.
53. K. Saito, K. Yonekura-Sakakibara, R. Nakabayashi, Y. Higashi, M. Yamazaki, T. Tohge and A. R. Fernie, *Plant Physiology and Biochemistry*, 2013, **72**, 21-34.
54. B. Winkel-Shirley, in *Advances in Botanical Research*, 2002, pp. 75-94.
55. L. Lepiniec, I. Debeaujon, J.-M. Routaboul, A. Baudry, L. Pourcel, N. Nesi and M. Caboche, *Annual Review of Plant Biology*, 2006, **57**, 405-430.
56. P. Jones, B. Messner, J. Nakajima, A. R. Schaffner and K. Saito, *The Journal of biological chemistry*, 2003, **278**, 43910-43918.
57. T. Tohge, Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., Saito, K., *Pure Appl Chem*, 2007, **79**, 811-823.
58. T. Tohge, Y. Nishiyama, M. Y. Hirai, M. Yano, J. Nakajima, M. Awazuhara, E. Inoue, H. Takahashi, D. B. Goodenowe, M. Kitayama, M. Noji, M. Yamazaki and K. Saito, *The Plant journal : for cell and molecular biology*, 2005, **42**, 218-235.
59. K. Yonekura-Sakakibara, A. Fukushima, R. Nakabayashi, K. Hanada, F. Matsuda, S. Sugawara, E. Inoue, T. Kuromori, T. Ito and et al., *Plant Journal*, 2012, **69**, 154-167.
60. K. Yonekura-Sakakibara, T. Tohge, F. Matsuda, R. Nakabayashi, H. Takayama, R. Niida, A. Watanabe-Takahashi, E. Inoue and K. Saito, *Plant Cell*, 2008, **20**, 2160-2176.
61. K. Yonekura-Sakakibara, T. Tohge, R. Niida and K. Saito, *The Journal of biological chemistry*, 2007, **282**, 14932-14941.
62. K. Yonekura-Sakakibara, R. Nakabayashi, S. Sugawara, T. Tohge, T. Ito, M. Koyanagi, M. Kitajima, H. Takayama and K. Saito, *The Plant journal : for cell and molecular biology*, 2014, **79**, 769-782.
63. J. Luo, Y. Nishiyama, C. Fuell, G. Taguchi, K. Elliott, L. Hill, Y. Tanaka, M. Kitayama, M. Yamazaki, P. Bailey, A. Parr, A. J. Michael, K. Saito and C. Martin, *The Plant journal : for cell and molecular biology*, 2007, **50**, 678-695.
64. J. C. D'Auria, M. Reichelt, K. Luck, A. Svatos and J. Gershenzon, *FEBS Letters*, 2007, **581**, 872-878.
65. I. Muzac, J. Wang, D. Anzellotti, H. Zhang and R. K. Ibrahim, *Archives of biochemistry and biophysics*, 2000, **375**, 385-388.
66. F. Matsuda, M. Y. Hirai, E. Sasaki, K. Akiyama, K. Yonekura-Sakakibara, N. J. Provart, T. Sakurai, Y. Shimada and K. Saito, *Plant physiology*, 2010, **152**, 566-578.
67. F. Matsuda, R. Nakabayashi, Y. Sawada, M. Suzuki, M. Y. Hirai, S. Kanaya and K. Saito, *Frontiers in plant science*, 2011, **2**, 40.
68. A. Fukushima, M. Kusano, R. F. Mejia, M. Iwasa, M. Kobayashi, N. Hayashi, A. Watanabe-Takahashi, T. Narisawa, T. Tohge, M. Hur, E. S. Wurtele, B. J. Nikolau and K. Saito, *Plant physiology*, 2014, **165**, 948-961.
69. F. Matsuda, Y. Okazaki, A. Oikawa, M. Kusano, R. Nakabayashi, J. Kikuchi, J.-I. Yonemaru, K. Ebana, M. Yano and et al., *Plant Journal*, 2012, **70**, 624-636.
70. L. Gong, W. Chen, Y. Gao, X. Liu, H. Zhang, C. Xu, S. Yu, Q. Zhang and J. Luo, *Proceedings of the National Academy of Sciences*, 2013.
71. M. Brazier-Hicks, K. M. Evans, M. C. Gershater, H. Puschmann, P. G. Steel and R. Edwards, *Journal of Biological Chemistry*, 2009, **284**, 17926-17934.
72. R. A. Dixon, D.-Y. Xie and S. B. Sharma, *New Phytologist*, 2005, **165**, 9-28.
73. D.-Y. Xie, S. B. Sharma, N. L. Paiva, D. Ferreira and R. A. Dixon, *Science*, 2003, **299**, 396-399.
74. Y. Pang, X. Cheng, D. Huhman, J. Ma, G. Peel, K. Yonekura-Sakakibara, K. Saito, G. Shen, L. Sumner, Y. Tang, J. Wen, J. Yun and R. Dixon, *Planta*, 2013, **238**, 139-154.
75. S. Kitamura, F. Matsuda, T. Tohge, K. Yonekura-Sakakibara, M. Yamazaki, K. Saito and I. Narumi, *The Plant journal : for cell and molecular biology*, 2010, **62**, 549-559.
76. L. Pourcel, J.-M. Routaboul, L. Kerhoas, M. Caboche, L. Lepiniec and I. Debeaujon, *The Plant Cell Online*, 2005, **17**, 2966-2980.
77. J. Zhao, Y. Pang and R. A. Dixon, *Plant physiology*, 2010, **153**, 437-443.
78. M. Y. Hirai, M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara and K. Saito, *Proc Natl Acad Sci U S A*, 2004, **101**, 10205-10210.
79. M. Y. Hirai, M. Klein, Y. Fujikawa, M. Yano, D. B. Goodenowe, Y. Yamazaki, S. Kanaya, Y. Nakamura, M. Kitayama and et al., *Journal of Biological Chemistry*, 2005, **280**, 25590-25595.
80. M. Y. Hirai, K. Sugiyama, Y. Sawada, T. Tohge, T. Obayashi, A. Suzuki, R. Araki, N. Sakurai, H. Suzuki, K. Aoki, H. Goda, O. I. Nishizawa, D. Shibata and K. Saito, *Proc Natl Acad Sci U S A*, 2007, **104**, 6478-6483.
81. T. Obayashi, K. Kinoshita, K. Nakai, M. Shibaoka, S. Hayashi, M. Saeki, D. Shibata, K. Saito and H. Ohta, *Nucleic Acids Research*, 2007, **35**, D863-D869.
82. E. K. F. Chan, H. C. Rowe, J. A. Corwin, B. Joseph and D. J. Kliebenstein, *PLoS Biol*, 2011, **9**, e1001125.
83. D. J. Kliebenstein, J. Kroymann, P. Brown, A. Figuth, D. Pedersen, J. Gershenzon and T. Mitchell-Olds, *Plant physiology*, 2001, **126**, 811-825.
84. B. G. Hansen, D. J. Kliebenstein and B. A. Halkier, *The Plant Journal*, 2007, **50**, 902-910.
85. F. Geu-Flores, M. E. Møldrup, C. Böttcher, C. E. Olsen, D. Scheel and B. A. Halkier, *Plant Cell*, 2011, **23**, 2456-2469.
86. J. M. Augustin, V. Kuzina, S. B. Andersen and S. Bak, *Phytochemistry*, 2011, **72**, 435-457.
87. S. Sawai and K. Saito, *Frontiers in plant science*, 2011, **2**.
88. A. Gholami, N. De Geyter, J. Pollier, S. Goormachtig and A. Goossens, *Natural product reports*, 2014, **31**, 356-380.

89. H. Seki, K. Ohyama, S. Sawai, M. Mizutani, T. Ohnishi, H. Sudo, T. Akashi, T. Aoki, K. Saito and et al., *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 14204-14209.
90. H. Seki, S. Sawai, K. Ohyama, M. Mizutani, T. Ohnishi, H. Sudo, E. O. Fukushima, T. Akashi, T. Aoki, K. Saito and T. Muranaka, *Plant Cell*, 2011, **23**, 4112-4123.
91. M. Carelli, E. Biazzi, F. Panara, A. Tava, L. Scaramelli, A. Porceddu, N. Graham, M. Odoardi, E. Piano, S. Arcioni, S. May, C. Scotti and O. Calderini, *The Plant Cell Online*, 2011, **23**, 3070-3081.
92. E. O. Fukushima, H. Seki, K. Ohyama, E. Ono, N. Umemoto, M. Mizutani, K. Saito and T. Muranaka, *Plant and Cell Physiology*, 2011, **52**, 2050-2061.
93. M. A. Naoumkina, L. V. Modolo, D. V. Huhman, E. Urbanczyk-Wochniak, Y. Tang, L. W. Sumner and R. A. Dixon, *Plant Cell*, 2010, **22**, 850-866.
94. J. M. Augustin, S. Drok, T. Shinoda, K. Sanmiya, J. K. Nielsen, B. Khakimov, C. E. Olsen, E. H. Hansen, V. Kuzina, C. T. Ekstrøm, T. Hauser and S. Bak, *Plant physiology*, 2012, **160**, 1881-1895.
95. T. Sayama, E. Ono, K. Takagi, Y. Takada, M. Horikawa, Y. Nakamoto, A. Hirose, H. Sasama, M. Ohashi, H. Hasegawa, T. Terakawa, A. Kikuchi, S. Kato, N. Tatsuzaki, C. Tsukamoto and M. Ishimoto, *Plant Cell*, 2012, **24**, 2123-2138.
96. A. Osbourn, *Plant physiology*, 2010, **154**, 531-535.
97. S. Boycheva, L. Daviet, J. L. Wolfender and T. B. Fitzpatrick, *Trends in plant science*, 2014.
98. X. Qi, S. Bakht, M. Leggett, C. Maxwell, R. Melton and A. Osbourn, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 8233-8238.
99. B. Field and A. E. Osbourn, *Science*, 2008, **320**, 543-547.
100. B. Field, A.-S. Fiston-Lavier, A. Kemen, K. Geisler, H. Quesneville and A. E. Osbourn, *Proceedings of the National Academy of Sciences*, 2011, **108**, 16116-16121.
101. M. Itkin, U. Heinig, O. Tzfadia, A. J. Bhide, B. Shinde, P. D. Cardenas, S. E. Bocobza, T. Unger, S. Malitsky, R. Finkers, Y. Tikunov, A. Bovy, Y. Chikate, P. Singh, I. Rogachev, J. Beekwilder, A. P. Giri and A. Aharoni, *Science*, 2013, **341**, 175-179.
102. S. Sawai, K. Ohyama, S. Yasumoto, H. Seki, T. Sakuma, T. Yamamoto, Y. Takebayashi, M. Kojima, H. Sakakibara, T. Aoki, T. Muranaka, K. Saito and N. Umemoto, *Plant Cell*, 2014.
103. A. Gesell, M. Rolf, J. Ziegler, M. L. Diaz Chavez, F. C. Huang and T. M. Kutchan, *The Journal of biological chemistry*, 2009, **284**, 24432-24442.
104. J. M. Hagel and P. J. Facchini, *Nature chemical biology*, 2010, **6**, 273-275.
105. T. Winzer, V. Gazda, Z. He, F. Kaminski, M. Kern, T. R. Larson, Y. Li, F. Meade, R. Teodor, F. E. Vaistij, C. Walker, T. A. Bowser and I. A. Graham, *Science*, 2012, **336**, 1704-1708.
106. F. Geu-Flores, N. H. Sherden, V. Courdavault, V. Burlat, W. S. Glenn, C. Wu, E. Nims, Y. Cui and S. E. O'Connor, *Nature*, 2012, **492**, 138-142.
107. K. Asada, V. Salim, S. Masada-Atsumi, E. Edmunds, M. Nagatoshi, K. Terasaka, H. Mizukami and V. De Luca, *The Plant Cell Online*, 2013, **25**, 4123-4134.
108. K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. van der Krol, R. Lugan, T. Ilc, R. Verpoorte, K. M. Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink and D. Werck-Reichhart, *Nature communications*, 2014, **5**, 3606.
109. V. Salim, B. Wiens, S. Masada-Atsumi, F. Yu and V. De Luca, *Phytochemistry*, 2014, **101**, 23-31.
110. M. Yamazaki, K. Mochida, T. Asano, R. Nakabayashi, M. Chiba, N. Udomson, Y. Yamazaki, D. Goodenowe, U. Sankawa, T. Yoshida, A. Toyoda, Y. Totoki, Y. Sakaki, E. Góngora-Castillo, C. Buell, T. Sakurai and K. Saito, *Plant & cell physiology*, 2013, **54**, 686-696.
111. T. Asano, K. Kobayashi, E. Kashihara, H. Sudo, R. Sasaki, Y. Iijima, K. Aoki, D. Shibata, K. Saito and M. Yamazaki, *Phytochemistry*, 2013, **91**, 128-139.
112. R. Nakabayashi and K. Saito, *Analytical and Bioanalytical Chemistry*, 2013, **405**, 5005-5011.
113. S. Bunsupa, M. Yamazaki and K. Saito, *Frontiers in plant science*, 2012, **3**, 239.
114. S. Bunsupa, K. Katayama, E. Ikeura, A. Oikawa, K. Toyooka, K. Saito and M. Yamazaki, *Plant Cell*, 2012, **24**, 1202-1216.
115. R. N. Stewart, K. H. Norris and S. Asen, *Phytochemistry*, 1975, **14**, 937-942.
116. G. Hrazdina and G. J. Wagner, in *The Biochemistry of Plant Phenols*, eds. C. F. Van Sumere and P. J. Lea, Oxford University Press, Oxford, England, 1985, pp. 119-133.
117. A. M. Zobel and G. Hrazdina, *Biotech Histochem*, 1995, **70**, 1-6.
118. M. E. Galway, J. D. Masucci, A. M. Lloyd, V. Walbot, R. W. Davis and J. W. Schiefelbein, *Developmental biology*, 1994, **166**, 740-754.
119. K. Noda, B. J. Glover, P. Linstead and C. Martin, *Nature*, 1994, **369**, 661-664.
120. W. H. Outlaw and S. Zhang, *Journal of experimental botany*, 2001, **52**, 605-614.
121. W. Wirtz, M. Stitt and H. W. Heldt, *Plant physiology*, 1980, **66**, 187-193.
122. M. Stitt, W. Wirtz and H. W. Heldt, *Biochimica et biophysica acta*, 1980, **593**, 85-102.
123. S. Klie, S. Krueger, L. Krall, P. Giavalisco, U. I. Flugge, L. Willmitzer and D. Steinhauser, *Frontiers in plant science*, 2011, **2**, 55.
124. J. Kehr, *Current opinion in plant biology*, 2003, **6**, 617-621.
125. M. L. Reyzer, Y. S. Hsieh, K. Ng, W. A. Korfmacher and R. M. Caprioli, *Journal of Mass Spectrometry*, 2003, **38**, 1081-1092.
126. R. L. Caldwell and R. M. Caprioli, *Molecular & Cellular Proteomics*, 2005, **4**, 394-401.
127. Y. J. Lee, D. C. Perdian, Z. H. Song, E. S. Yeung and B. J. Nikolau, *Plant Journal*, 2012, **70**, 81-95.
128. P. Chaurand, M. Stoeckli and R. Caprioli, *Anal Chem*, 1999, **71**, 5263-5270.
129. N. Ishida, M. Koizumi and H. Kano, *Ann. Bot.*, 2000, **86**, 259-278.
130. W. Kockenberger, *Trends in plant science*, 2001, **6**, 286-292.
131. L. Borisjuk, H. Rolletschek and T. Neuberger, *Plant Journal*, 2012, **70**, 129-146.
132. H. Van As and J. van Duynhoven, *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 2013, **229**, 25-34.
133. J. S. Fletcher, J. C. Vickerman and N. Winograd, *Current opinion in chemical biology*, 2011, **15**, 733-740.
134. E. H. Seeley and R. M. Caprioli, *Anal Chem*, 2012, **84**, 2105-2110.
135. J. S. Fletcher, N. P. Lockyer and J. C. Vickerman, *Mass spectrometry reviews*, 2011, **30**, 142-174.
136. S. Cha, H. Zhang, H. I. Ilarslan, E. S. Wurtele, L. Brachova, B. J. Nikolau and E. S. Yeung, *The Plant journal : for cell and molecular biology*, 2008, **55**, 348-360.
137. Y. J. Lee, D. C. Perdian, Z. Song, E. S. Yeung and B. J. Nikolau, *The Plant journal : for cell and molecular biology*, 2012, **70**, 81-95.
138. M. M. Burrell, C. J. Earnshaw and M. R. Clench, *Journal of Experimental Botany*, 2007, **58**, 757-763.
139. A. Benninghoven, F. G. Rudenauer and H. W. Werner, *Secondary ion mass spectrometry: basic concepts, instrumental aspects, applications and trends*, John Wiley & Sons, New York, 1987.
140. M. Pacholski and N. Winograd, *Chemical reviews*, 1999, **99**, 2977-3006.
141. D. B. Lazof, J. G. Goldsmith, T. W. Rufty and R. W. Linton, *Plant physiology*, 1996, **112**, 1289-1300.
142. R. Metzner, H. U. Schneider, U. Breuer and W. H. Schroeder, *Plant physiology*, 2008, **147**, 1774-1787.
143. K. L. Moore, M. Schroder, Z. Wu, B. G. H. Martin, C. R. Hawes, S. P. McGrath, M. J. Hawkesford, J. Feng Ma, F.-J. Zhao and C. R. M. Grovenor, *Plant physiology*, 2011, **156**, 913-924.
144. A. Seyer, J. Einhorn, A. Brunelle and O. Laprevote, *Anal Chem*, 2010, **82**, 2326-2333.
145. K. L. Moore, E. Lombi, F. J. Zhao and C. R. Grovenor, *Anal Bioanal Chem*, 2012, **402**, 3263-3273.

146. D. Hölscher, R. Shroff, K. Knop, M. Gottschaldt, A. Crecelius, B. Schneider, D. G. Heckel, U. S. Schubert and A. Svatoš, *The Plant journal : for cell and molecular biology*, 2009, **60**, 907-918.
147. S. Cha, Z. Song, B. J. Nikolau and E. S. Yeung, *Anal Chem*, 2009, **81**, 2991-3000.
148. P. Nemes, A. A. Barton, Y. Li and A. Vertes, *Analytical Chemistry*, 2008, **80**, 4575-4582.
149. P. Nemes, A. A. Barton and A. Vertes, *Analytical Chemistry*, 2009, **81**, 6668-6675.
150. P. Nemes and A. Vertes, *Analytical Chemistry*, 2007, **79**, 8098-8106.
151. P. Nemes, A. S. Woods and A. Vertes, *Analytical chemistry*, 2010, **82**, 982-988.
152. B. Shrestha, P. Nemes, J. Nazarian, Y. Hathout, E. P. Hoffman and A. Vertes, *The Analyst*, 2010, **135**, 751-758.
153. B. Shrestha, P. Nemes and A. Vertes, *Applied Physics A*, 2010, **101**, 121-126.
154. B. Shrestha, J. M. Patt and A. Vertes, *Analytical Chemistry*, 2011, **83**, 2947-2955.
155. A. Svatos, *Trends Biotechnol*, 2010, **28**, 425-434.
156. M. Peukert, A. Matros, G. Lattanzio, S. Kaspar, J. Abadia and H. P. Mock, *The New phytologist*, 2012, **193**, 806-815.
157. T. Mueller, S. Oradu, D. R. Ifa, R. G. Cooks and B. Kraeutler, *Analytical Chemistry*, 2011, **83**, 5754-5761.
158. L. W. Sumner, D. S. Yang, B. J. Bench, B. S. Watson, C. Li and A. D. Jones, *Annual Plant Reviews*, 2011, **43**, 343-366.
159. D. C. Perdian and Y. J. Lee, *Anal Chem*, 2010, **82**, 9393-9400.
160. A. K. Mullen, M. R. Clench, S. Crosland and K. R. Sharples, *Rapid communications in mass spectrometry : RCM*, 2005, **19**, 2507-2516.
161. S. Robinson, K. Warburton, M. Seymour, M. Clench and J. Thomas-Oates, *The New phytologist*, 2007, **173**, 438-444.
162. V. Vrkoslav, A. Muck, J. Cvacka and A. Svatos, *Journal of the American Society for Mass Spectrometry*, 2010, **21**, 220-231.
163. P. Franceschi, Y. Dong, K. Strupat, U. Vrhovsek and F. Mattivi, *J Exp Bot*, 2012, **63**, 1123-1133.
164. M. Ha, J. H. Kwak, Y. Kim and O. P. Zee, *Food Chemistry*, 2012, **133**, 1155-1162.
165. J. Sarsby, M. W. Towers, C. Stain, R. Cramer and O. A. Koroleva, *Phytochemistry*, 2012, **77**, 110-118.
166. Y. Yoshimura, H. Enomoto, T. Moriyama, Y. Kawamura, M. Setou and N. Zaima, *Anal Bioanal Chem*, 2012, **403**, 1885-1895.
167. Y. Yoshimura, N. Zaima, T. Moriyama and Y. Kawamura, *PloS one*, 2012, **7**, e31285.
168. H. Ye, E. Gemperline, M. Venkateshwaran, R. Chen, P. M. Delaux, M. Howes-Podoll, J. M. Ane and L. Li, *The Plant journal : for cell and molecular biology*, 2013, **75**, 130-145.
169. J. H. Jun, Z. H. Song, Z. J. Liu, B. J. Nikolau, E. S. Yeung and Y. J. Lee, *Analytical Chemistry*, 2010, **82**, 3255-3265.
170. S. Cha, Z. Song, B. J. Nikolau and E. S. Yeung, *Analytical Chemistry*, 2009, **81**, 2991-3000.
171. C. Sluszny, E. S. Yeung and B. J. Nikolau, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 107-115.
172. Y. J. Lee, D. C. Perdian, Z. Song, E. S. Yeung and B. J. Nikolau, *The Plant journal : for cell and molecular biology*, 2012, **70**, 81-95.
173. R. G. Cooks, Z. Ouyang, Z. Takats and J. M. Wiseman, *Science*, 2006, **311**, 1566-1570.
174. A. B. Costa and R. Graham Cooks, *Chemical Physics Letters*, 2008, **464**, 1-8.
175. I. Cotte-Rodriguez, C. C. Mulligan and R. G. Cooks, *Anal Chem*, 2007, **79**, 7069-7077.
176. Q. Hu, N. Talaty, R. J. Noll and R. G. Cooks, *Rapid communications in mass spectrometry*, 2006, **20**, 3403-3408.
177. R. G. Cooks, N. E. Manicke, A. L. Dill, D. R. Ifa, L. S. Eberlin, A. B. Costa, H. Wang, G. Huang and Z. Ouyang, *Faraday discussions*, 2011, **149**, 247-267.
178. A. L. Dill, D. R. Ifa, N. E. Manicke, A. B. Costa, J. A. Ramos-Vara, D. W. Knapp and R. G. Cooks, *Analytical chemistry*, 2009, **81**, 8758-8764.
179. L. S. Eberlin, A. L. Dill, A. B. Costa, D. R. Ifa, L. Cheng, T. Masterson, M. Koch, T. L. Ratliff and R. G. Cooks, *Analytical chemistry*, 2010, **82**, 3430-3434.
180. S. R. Ellis, C. Wu, J. M. Deeley, X. Zhu, R. J. Truscott, R. G. Cooks, T. W. Mitchell and S. J. Blanksby, *Journal of the American Society for Mass Spectrometry*, 2010, **21**, 2095-2104.
181. M. Girod, Y. Shi, J.-X. Cheng and R. G. Cooks, *Analytical chemistry*, 2010, **83**, 207-215.
182. D. R. Ifa, J. M. Wiseman, Q. Song and R. G. Cooks, *International Journal of Mass Spectrometry*, 2007, **259**, 8-15.
183. N. E. Manicke, M. Nefliu, C. Wu, J. W. Woods, V. Reiser, R. C. Hendrickson and R. G. Cooks, *Analytical chemistry*, 2009, **81**, 8702-8707.
184. J. M. Wiseman, D. R. Ifa, Y. Zhu, C. B. Kissinger, N. E. Manicke, P. T. Kissinger and R. G. Cooks, *Proceedings of the National Academy of Sciences*, 2008, **105**, 18120-18125.
185. A. U. Jackson, A. Tata, C. Wu, R. H. Perry, G. Haas, L. West and R. G. Cooks, *The Analyst*, 2009, **134**, 867-874.
186. N. Talaty, Z. Takats and R. G. Cooks, *The Analyst*, 2005, **130**, 1624-1633.
187. P. J. Roach, J. Laskin and A. Laskin, *The Analyst*, 2010, **135**, 2233-2236.
188. A. R. Korte, Z. H. Song, B. J. Nikolau and Y. J. Lee, *Analytical Methods*, 2012, **4**, 474-481.
189. M. Koestler, D. Kirsch, A. Hester, A. Leisner, S. Guenther and B. Spengler, *Rapid communications in mass spectrometry : RCM*, 2008, **22**, 3275-3285.
190. A. Römpp, S. Guenther, Y. Schober, O. Schulz, Z. Takats, W. Kummer and B. Spengler, *Angewandte chemie international edition*, 2010, **49**, 3834-3838.
191. P. C. Lauterbur, *Nature*, 1973, **242**, 190-191.
192. M. Rokitta, A. Peuke, U. Zimmermann and A. Haase, *Protoplasma*, 1999, **209**, 126-131.
193. T. W. J. Scheenen, D. van Dusschoten, P. A. de Jager and H. Van As, *J. Exp. Bot.*, 2000, **51**, 1751-1759.
194. T. W. J. Scheenen, F. J. Vergeldt, A. M. Heemskerk and H. Van As, *Plant physiology*, 2007, **144**, 1157-1165.
195. H. Van As, *J Exp Bot*, 2007, **58**, 743-756.
196. C. W. Windt, E. Gerkema and H. Van As, *Plant physiology*, 2009, **151**, 830-842.
197. L. Ciobanu, D. A. Seeber and C. H. Pennington, *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 2002, **158**, 178-182.
198. G. Melkus, H. Rolletschek, J. Fuchs, V. Radchuk, E. Grafahrend-Belau, N. Sreenivasulu, T. Rutten, D. Weier, N. Heinzel, F. Schreiber, T. Altmann, P. M. Jakob and L. Borisjuk, *Plant biotechnology journal*, 2011, **9**, 1022-1037.
199. G. Melkus, H. Rolletschek, R. Radchuk, J. Fuchs, T. Rutten, U. Wobus, T. Altmann, P. Jakob and L. Borisjuk, *Plant physiology*, 2009, **151**, 1139-1154.
200. J. Lambert, P. Lampen, A. von Bohlen and R. Hergenroder, *Anal Bioanal Chem*, 2006, **384**, 231-236.
201. A. Rollins, J. Barber, R. Elliott and B. Wood, *Plant physiology*, 1989, **91**, 1243-1246.
202. L. Borisjuk, H. Rolletschek and T. Neuberger, *Progress in lipid research*, 2013, **52**, 465-487.
203. J. Fuchs, T. Neuberger, H. Rolletschek, S. Schiebold, T. H. Nguyen, N. Borisjuk, A. Borner, G. Melkus, P. Jakob and L. Borisjuk, *Plant physiology*, 2013, **161**, 583-593.
204. R. J. Bino, R. D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B. J. Nikolau, P. Mendes, U. Roessner-Tunali, M. H. Beale, R. N. Trethewey, B. M. Lange, E. S. Wurtele and L. W. Sumner, *Trends Plant Sci.*, 2004, **9**, 418-425.
205. R. M. Salek, C. Steinbeck, M. R. Viant, R. Goodacre and W. B. Dunn, *GigaScience*, 2013, **2**, 13.
206. W. B. Dunn, A. Erban, R. J. M. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann and J. Kopka, *Metabolomics : Official journal of the Metabolomic Society*, 2013, **9**, S44-S66.
207. L. Sumner, A. Amberg, D. Barrett, M. Beale, R. Beger, C. Daykin, T. Fan, O. Fiehn, R. Goodacre, J. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. Lane, J. Lindon, P. Marriott, A. Nicholls, M. Reily, J. Thaden and M. Viant, *Metabolomics*, 2007, **3**, 211-221.
208. J. L. Wolfender, E. F. Queiroz and K. Hostettmann, *Magn Reson Chem*, 2005, **43**, 697-709.

209. M. Lambert, J. L. Wolfender, D. Staerk, S. B. Christensen, K. Hostettmann and J. W. Jaroszewski, *Anal Chem*, 2007, **79**, 727-735.
210. A. Castro, S. Moco, J. Coll and J. Vervoort, *J Nat Prod*, 2010, **73**, 962-965.
211. J. van der Hooft, V. Mihaleva, R. de Vos, R. Bino and J. Vervoort, *Magn Reson Chem*, 2011, **49 Suppl 1**, S55-60.
212. H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *Journal of mass spectrometry : JMS*, 2010, **45**, 703-714.
213. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic acids research*, 2013, **41**, D801-807.
214. C. A. Smith, G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak, *Ther Drug Monit*, 2005, **27**, 747-751.
215. R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G. J. Patti and G. Siuzdak, *Nature biotechnology*, 2012, **30**, 826-828.
216. J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie and D. Steinhauser, *Bioinformatics (Oxford, England)*, 2005, **21**, 1635-1638.
217. Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M. Y. Hirai and K. Saito, *Phytochemistry*, 2012, **82**, 38-45.
218. F. Matsuda, K. Yonekura-Sakakibara, R. Niida, T. Kuromori, K. Shinozaki and K. Saito, *The Plant journal : for cell and molecular biology*, 2009, **57**, 555-577.
219. H. Neuweger, S. P. Albaum, M. Dondrup, M. Persicke, T. Watt, K. Niehaus, J. Stoye and A. Goesmann, *Bioinformatics (Oxford, England)*, 2008, **24**, 2726-2732.
220. Q. Cui, I. A. Lewis, A. D. Hegeman, M. E. Anderson, J. Li, C. F. Schulte, W. M. Westler, H. R. Eghbalnia, M. R. Sussman and J. L. Markley, *Nat Biotechnol*, 2008, **26**, 162-164.
221. E. Champarnaud and C. Hopley, *Rapid communications in mass spectrometry : RCM*, 2011, **25**, 1001-1007.
222. K. Scheubert, F. Hufsky and S. Bocker, *J Cheminform*, 2013, **5**, 12.
223. M. Gerlich and S. Neumann, *Journal of mass spectrometry : JMS*, 2013, **48**, 291-298.
224. S. Wolf, S. Schmidt, M. Muller-Hannemann and S. Neumann, *BMC bioinformatics*, 2010, **11**, 148.
225. M. Heinonen, A. Rantanen, T. Mielikainen, J. Kokkonen, J. Kiuru, R. A. Ketola and J. Rousu, *Rapid communications in mass spectrometry : RCM*, 2008, **22**, 3043-3052.
226. T. Kind, K. H. Liu, Y. Lee do, B. DeFelice, J. K. Meissen and O. Fiehn, *Nat Methods*, 2013, **10**, 755-758.
227. F. Rasche, K. Scheubert, F. Hufsky, T. Zichner, M. Kai, A. Svatos and S. Bocker, *Analytical chemistry*, 2012, **84**, 3417-3426.
228. S. E. Stein, *Journal of the American Society for Mass Spectrometry*, 1995, **6**, 644-655.
229. M. L. Bandu, K. R. Watkins, M. L. Bretthauer, C. A. Moore and H. Desaire, *Analytical chemistry*, 2004, **76**, 1746-1753.
230. K. Klagkou, F. Pullen, M. Harrison, A. Organ, A. Firth and G. J. Langley, *Rapid communications in mass spectrometry : RCM*, 2003, **17**, 1163-1168.
231. J. A. Taylor and R. S. Johnson, *Rapid communications in mass spectrometry : RCM*, 1997, **11**, 1067-1075.
232. L. J. Kangas, T. O. Metz, G. Isaac, B. T. Schrom, B. Ginovska-Pangovska, L. Wang, L. Tan, R. R. Lewis and J. H. Miller, *Bioinformatics (Oxford, England)*, 2012, **28**, 1705-1713.
233. H. Zhang, S. Singh and V. N. Reinhold, *Analytical chemistry*, 2005, **77**, 6263-6270.
234. A. Kameyama, S. Nakaya, H. Ito, N. Kikuchi, T. Angata, M. Nakamura, H. K. Ishida and H. Narimatsu, *J Proteome Res*, 2006, **5**, 808-814.
235. R. Wei, G. Li and A. B. Seymour, *Analytical chemistry*, 2010, **82**, 5527-5533.
236. Y. Sawada, K. Akiyama, A. Sakata, A. Kuwahara, H. Otsuki, T. Sakurai, K. Saito and M. Y. Hirai, *Plant & cell physiology*, 2009, **50**, 37-47.
237. U. Vrhovsek, D. Masuero, M. Gasperotti, P. Franceschi, L. Caputi, R. Viola and F. Mattivi, *Journal of agricultural and food chemistry*, 2012, **60**, 8831-8840.
238. E. Carvalho, P. Franceschi, A. Feller, L. Palmieri, R. Wehrens and S. Martens, *Plant Physiol Biochem*, 2013, **72**, 79-86.
239. A. Schurmann, V. Dvorak, C. Cruzer, P. Butcher and A. Kaufmann, *Rapid Commun Mass Spectrom*, 2009, **23**, 1196-1200.
240. D. Hurtaud-Pessel, E. Verdon, J. Blot and P. Sanders, *Food Addit Contam*, 2006, **23**, 569-578.
241. B. J. Berendsen, L. A. Stolker and M. W. Nielen, *Journal of the American Society for Mass Spectrometry*, 2013, **24**, 154-163.
242. Ó. J. Pozo, J. V. Sancho, M. Ibáñez, F. Hernández and W. M. A. Niessen, *TrAC Trends in Analytical Chemistry*, 2006, **25**, 1030-1042.
243. H. Gao, O. L. Materne, D. L. Howe and C. L. Brummel, *Rapid communications in mass spectrometry : RCM*, 2007, **21**, 3683-3693.
244. R. G. Dickinson and A. L. Raymond, *J. Amer. Chem. Soc.*, 1923, **45**, 22-29.
245. J. P. Glusker, *Protein Science*, 1994, **3**, 2465-2469.
246. Y. Inokuma, S. Yoshioka, J. Ariyoshi, T. Arai, Y. Hitora, K. Takada, S. Matsunaga, K. Rissanen and M. Fujita, *Nature*, 2013, **495**, 461-466.
247. Y. Inokuma, S. Yoshioka, J. Ariyoshi, T. Arai and M. Fujita, *Nature protocols*, 2014, **9**, 246-252.
248. N. Carreno-Quintero, H. J. Bouwmeester and J. J. B. Keurentjes, *Trends in Genetics*, 2013, **29**, 41-50.
249. L. Gong, W. Chen, Y. Gao, X. Liu, H. Zhang, C. Xu, S. Yu, Q. Zhang and J. Luo, *Proceedings of the National Academy of Sciences*, 2013, **110**, 20320-20325.
250. H. Redestig, M. Kusano, K. Ebana, M. Kobayashi, A. Oikawa, Y. Okazaki, F. Matsuda, M. Arita, N. Fujita and et al., *BMC Systems Biology*, 2011, **5**, 176
251. M. N. Calingacion, C. Boualaphanh, V. D. Daygon, R. Anacleto, R. Sackville Hamilton, B. Biais, C. Deborde, M. Maucourt, A. Moing, R. Mumm, R. C. H. Vos, A. Erban, J. Kopka, T. H. Hansen, K. H. Laursen, J. K. Schjoerring, R. D. Hall and M. A. Fitzgerald, *Metabolomics : Official journal of the Metabolomic Society*, 2012, **8**, 771-783.
252. F. Matsuda, R. Nakabayashi, Z. Yang, Y. Okazaki, J. Yonemaru, K. Ebana, M. Yano and K. Saito, *Plant Journal*, 2014, **accepted**.
253. Z. Yang, R. Nakabayashi, Y. Okazaki, T. Mori, S. Takamatsu, S. Kitanaka, J. Kikuchi and K. Saito, *Metabolomics : Official journal of the Metabolomic Society*, 2013, 1-13.
254. J. B. Bowne, T. A. Erwin, J. Juttner, T. Schnurbusch, P. Langridge, A. Bacic and U. Roessner, *Molecular Plant*, 2012, **5**, 418-429.
255. C. B. Hill, J. D. Taylor, J. Edwards, D. Mather, A. Bacic, P. Langridge and U. Roessner, *Plant physiology*, 2013, **162**, 1266-1281.
256. N. Amiour, S. Imbaud, G. Clément, N. Agier, M. Zivy, B. Valot, T. Balliau, P. Armengaud, I. Quilleré, R. Cañas, T. Tercet-Laforgue and B. Hirel, *Journal of Experimental Botany*, 2012, **63**, 5017-5033.
257. P. Casati, M. Campi, D. Morrow, J. Fernandes and V. Walbot, *BMC Genomics*, 2011, **12**, 321.
258. J. Qiu, *Nature*, 2007, **448**, 126-128.
259. M. Wang, R.-J. A. N. Lamers, H. A. A. J. Korthout, J. H. J. van Nesselrooij, R. F. Witkamp, R. van der Heijden, P. J. Voshol, L. M. Havekes, R. Verpoorte and J. van der Greef, *Phytotherapy Research*, 2005, **19**, 173-182.
260. A. Zhang, H. Sun, Z. Wang, W. Sun, P. Wang and X. Wang, *Planta Med*, 2010, **76**, 2026-2035.
261. T. Okada and e. al., *Submitted*, 2014.