DOCTORAL THESIS

Novel software for stable isotopic labelling assisted and LC-HRMS based untargeted metabolomics research and tracer-fate studies

Christoph Bueschl

Bioinformatics
Institute for Knowledge Discovery
Graz University of Technology
Petersgasse 14, A-8010 Graz

Advisors:
Rainer Schuhmacher
Gerhard G. Thallinger

Reviewers:
Rudolf Stollberger
Karsten Hiller
Rudolf Krska

Examiners:
Rudolf Stollberger
Karsten Hiller
Rainer Schuhmacher

Graz, December 2014

This page is left blank intentionally

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material, which has been quoted either literally or by content from the used sources. The text document uploaded to TUGRAZonline is identical to the present doctoral dissertation.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Dissertation identisch.

date (signature)

Acknowledgments

I would like to express my gratitude to the many people, who I had the pleasure working with and who supported me during my doctoral study. It was a privilege collaborating with each and every one of you.

I am utmost thankful to Rainer Schuhmacher, head of the Metabolomics team at the Center for Analytical Chemistry, IFA-Tulln. It was your enthusiasm that initiated our collaboration and I very much enjoyed discussing all project-related aspects with you. Thank you very much for your invaluable support and motivation.

I want to express my sincere gratitude to Bernhard Kluger, who helped me understand the many facets of analytical chemistry and metabolomics research. It was a privilege working with you and your support and motivating force shaped so many aspects of our work.

I am also especially grateful to Gerhard Thallinger, head of the Thallinger Lab at the Graz University of Technology. I very much enjoyed presenting and discussing our work with you and your feedback and ideas influenced many facets of my doctoral study.

I am very thankful to Nora Neumann, Sylvia Lehner, Benedikt Warth, Elisabeth Varga, Jürgen Hartler, Alexandra Parich, Denise Schöfbeck, Maria Doppler, David McMullin, Jacqueline Reiterer, Alexandra Simader, Oana Tomescu, Franz Berthiller, Renate Hörmann, and Rudolf Krska. Thank you for all your help, support, and feedback about our work.

I want to thank our collaborators Gerhard Adam, Josef Strauss, Stephan Bödi, Marc Lemmens, Valentina Maschietto, Adriano Marocco, Alexander Chassy, and Andrew Waterhouse for the exciting projects and your help in understanding the biological motivation.

I would like to thank all current and previous members of the Center for Analytical Chemistry and the Thallinger Lab. Your support, feedback and motivation illuminated the many days of my doctoral study.

I am also greatly thankful to my parents, Renate and Franz, my brother, Christian, and my grandmother, Leopoldine. Thank you for your restless encouragement, support, and for letting me known that I could always rely on you.

I also want to thank god. Besides all the questions, that will be and have been answered with science, there are so many more aspects and facets that further feed our curiosity.

Finally, I want to acknowledge support by the University of Natural Resources and Life Sciences, Vienna (BOKU) and the Graz University of Technology (TU Graz), as well as The Austrian Science Fund (FWF) and the Vienna Science and Technology Fund (WWTF).

Abstract

Motivation: LC-HRMS is the most commonly used analytical technique for the simultaneous detection of as many metabolites as possible in untargeted metabolomics approaches. Despite its many advantages, it also has several drawbacks including unspecific, non-biological related signals and matrix effects that distort relative metabolite abundances and complicate subsequent statistical analysis. In this respect, stable isotopic labelling, which relies on the artificial creation of metabolite molecules enriched with certain stable isotopes (e.g. ^{13}C or ^{15}N), can be used to overcome these limitations.

Aim: The aim of the presented thesis was the development of novel software for the processing of LC-HRMS data derived from native and isotopically labelled biological samples.

Results: The developed software MetExtract facilitates a) the automated detection of all metabolites of a biological system and b) the automated detection of all biotransformation products derived from a studied tracer substance in stable isotopic labelling assisted and LC-HRMS based untargeted metabolomics applications. The software utilises the distinct isotope patterns introduced by native and labelled isotopologs of the same metabolites for the detection of all truly metabolite-derived signals and thus efficiently removes any information of non-biological origin. Each detected feature pair that represents a certain ion species of a metabolite is verified with i) the distinct molecular mass gain introduced by the labelling, ii) the unique isotope patterns, and iii) a check for congruent coelution of the native and the labelled metabolite forms. Additionally, the software reports the exact number of incorporated labelling-isotope atoms for each found feature pair, which greatly enhances metabolite annotation. Moreover, MetExtract convolutes feature pairs derived from the same metabolite into feature groups and determines a fold quantification value, which, provided a proper labelling and sample pooling protocol ahead of LC-HRMS analysis, accounts for different ion suppression and enhancement effects as well as mass detector drifts across different measurement batches and improves relative metabolite quantification and thus comparison of experimental conditions.

Applications: With the aid of MetExtract the advantages of SIL in untargeted metabolomics approaches, which include the detection of only biological metabolites (e.g. 135 *Fusarium graminearum*- and 362 wheat-derived metabolites), improved relative metabolite quantification, and multivariate statistical analysis, were demonstrated.

In tracer-fate experiments MetExtract was used to automatically detect 9 detoxification products of the *F. graminearum* mycotoxin deoxynivalenol in wheat. Moreover, it was utilised for studying the metabolic fate of the aromatic amino acid phenylalanine in grape-berries (63 metabolites) and wheat cell suspension cultures (139 metabolites). All these metabolites were unambiguously assigned descendants of the respective tracer under investigation.

Keywords: MetExtract, Metabolite detection, Metabolite characterisation, Fast polarity switching

Kurzfassung

Motivation: Flüssigkeitschromatographie in Kombination mit hochauflösender Massenspektrometrie und Electrospray-Ionisierung (LC-ESI-HRMS) ist die derzeit wichtigste und am häufigsten eingesetzte analytische Methode um das Metabolome einer biologischen Probe zu erfassen. Trotz vieler Vorteile hat diese Technik die Nachteile, dass nicht zwischen biologisch relevanten Metaboliten und Kontaminationen unterschieden wird und dass Matrix-Effekte die relative Quantifizierung der erfassten Metabolite verzerren. Isotopenmarkierung mit stabilen Isotopen gleicht diese Nachteile aus und verbessert dadurch die Interpretation des Versuchs.
Ziele: Zur effizienten Nutzung der Isotopenmarkierung sollte im Zuge dieser Dissertation eine Software zur automatischen Auswertung von LC-HRMS Daten natürlicher und isotopenmarkierter biologischer Proben entwickelt werden.
Ergebnisse: Die Dissertation beschreibt die Software MetExtract, die mittels LC-HRMS Analyse generierte Metabolomics Daten analysiert. MetExtract detektiert a) alle Metabolite eines biologischen Systems und b) nur jene Metabolite, die von einer bestimmten Vorläufersubstanz abstammen. Hierfür bedient sich MetExtract komplexer Isotopenmuster, die durch das Markieren mittels stabiler Isotope in den jeweiligen biologischen Experimenten entstehen. Die Software filtert Störsignale und Substanzen nicht biologischen Ursprungs aus und erkennt daher nur biologisch relevanten Metabolite. Jeder gefundenen Metabolite wird mit 3 Kriterien verifiziert: i) existiert einen isotopenmarkierter Partner mit der entsprechenden Masse, ii) ist das Isotopenmuster gültig und iii) sind die natürliche und markierte Metabolitformen identisch chromatografiert. Anhand dieser Merkmale wird für jeden gefundenen Metabolit auch die Anzahl an ausgetauschten Atomen mit dem verwendeten stabilen Isotop ermittelt. Unterschiedliche Ionen, die vom selben Metabolit abstammen, werden in Metabolitgruppen zusammengefasst. Des Weiteren errechnet MetExtract einen Fold-quantification Wert, der, sofern das Experiment entsprechend durchgeführt wurde, unterschiedliche Matrix-effekte und Detekorabweichungen ausgleicht, wodurch die relativen Konzentrationen der Metabolite korrigiert und die statistische Auswertung des Experiments verbessert wird.
Anwendungen: MetExtract wurde angewendet, um die Vorteile von Isotopenmarkierung mittels stabiler Isotope in ungerichteten Metabolomicsexperimenten zu demonstrieren.
Des Weiteren wurde das entwickelte Tool eingesetzt, um den Metabolismus des Mycotoxins Deoxynivalenol in Weizenpflanzen zu untersuchen. Dieser Ansatz fand 9 Detoxifizierungsprodukte von Deoxynivalenol. Auf ähnliche Weise wurde auch nach sekundären Metaboliten, die von der Aminosäure Phenylalanine abstammen, in Weinbeeren (63 Metabolite) und Weizenzellkulturen (139 Metabolite) gesucht. Neben bereits bekannten wurden auch neue, unbekannte Metabolite in den Proben gefunden und annotiert.

Suchbegriffe: MetExtract, Metabolit Detektion, Metabolit Charakterisierung, Fast polarity switching

Publications

During the author's doctoral study, eight manuscripts were published in peer-reviewed journals in the field of untargeted metabolomics research. Five of these (numbered; reprinted in Part II) form an integral part of the presented doctoral thesis:

MetExtract: a new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research

Christoph Bueschl, Bernhard Kluger, Franz Berthiller, Gerald Lirk, Stefan Winkler, Rudolf Krska, Rainer Schuhmacher
Bioinformatics. 2012;28(5):736-738

#1 *Isotopic labelling-assisted metabolomics using LC-MS*

Christoph Bueschl, Rudolf Krska, Bernhard Kluger, Rainer Schuhmacher
Analytical and Bioanalytical Chemistry. 2013;405(1):27-33

#2 *Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat*

Bernhard Kluger, **Christoph Bueschl**, Marc Lemmens, Franz Berthiller, Georg Häubl, Günther Jaunecker, Gerhard Adam, Rudolf Krska, Rainer Schuhmacher
Analytical and Bioanalytical Chemistry. 2013;405(15):5031-5036

#3 *A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research*

Christoph Bueschl†, Bernhard Kluger†, Marc Lemmens, Gerhard Adam, Gerlinde Wiesenberger, Valentina Maschietto, Adriano Marocco, Joseph Strauss, Stephan Bödi, Gerhard G. Thallinger, Rudolf Krska, Rainer Schuhmacher
Metabolomics. 2014;10:754-769

#4 *Tracing flavonoid degradation in grapes by MS filtering with stable isotopes*

Alexander W. Chassy, **Christoph Bueschl**, Hyeyoung Lee, Larry Lerno, Anita Oberholster, Daniela Barile, Rainer Schuhmacher, Andrew L. Waterhouse
Food Chemistry. 2015;166:448-455

† Authors have contributed equally to this work

Automated LC-HRMS (/MS) approach for the annotation of fragment ions derived from stable isotope labelling-assisted untargeted metabolomics

Nora KN. Neumann, Sylvia M. Lehner, Bernhard Kluger, **Christoph Bueschl**, Karoline Sedelmaier, Marc Lemmens, Rudolf Krska, Rainer Schuhmacher
Analytical Chemistry. 2014;86(15):7320-7327

GC-MS based targeted metabolic profiling identifies changes in the wheat metabolome following deoxynivalenol treatment

Benedikt Warth, Alexandra Parich, **Christoph Bueschl**, Denise Schoefbeck, Nora KN. Neumann, Bernhard Kluger, Katharina Schuster, Rudolf Krska, Gerhard Adam, Marc Lemmens, Rainer Schuhmacher
Metabolomics. 2014; doi: 10.1007/s11306-014-0731-1

#5 *Untargeted profiling of tracer derived metabolites using stable isotopic labelling and fast polarity switching LC-ESI-HRMS*

Bernhard Kluger†, **Christoph Bueschl**†, Nora KN. Neumann, Romana Stückler, Maria Doppler, Alexander Chassy, Andrew Waterhouse, Justyna Rechthaler, Niklas Kampleitner, Gerhard G. Thallinger, Gerhard Adam, Rudolf Krska, Rainer Schuhmacher
Analytical Chemistry. 2014;86(23):11533-11537

† Authors have contributed equally to this work

Table of contents

Part I

Introduction

1. Scientific background

"We can only see a short distance ahead, but we can see plenty there that needs to be done"

- Alan Turing
(Computing Machinery and Intelligence, 1950, p. 460)

Wolfram Weckwerth, a pioneer in the young field of metabolomics research, states: "*The primary aim of 'omic' technologies is the nontargeted identification of all gene products (transcripts, proteins, and metabolites) present in a specific biological sample*" [6].
Metabolomics research, in this regard, focuses on the comprehensive study of low-molecular weight substances present in a biological system (e.g. an entire organism, a certain tissue type, or individual cells). Such substances are named metabolites and the entire set of all metabolites in a particular, biological system is referred to as its metabolome [7].

Oliver Fiehn, another topmost recognised pioneer, who has been working in the field of metabolomics research since its early days, writes: "*Metabolites are the end products of cellular regulatory processes, and their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes*" [8].
This stresses that the phenotype of a biological system is highly correlated with its metabolome as a means of functional genomics [9]. Thus, metabolomics research can be used for numerous applications: For example, a metabolite, which is easily quantified, may serve as a biomarker of a genetic condition or an environmental impact [10, 11]. Other metabolomics applications are concerned with or are used in support for the detection of novel therapeutic agents for various diseases [12-16]. Furthermore, metabolomics is applied in cancer research [17-19], comparison of different genotypes [20], toxicology [21], and nutrition [22, 23].

Metabolites commonly constitute several atoms of the elements carbon, hydrogen, and oxygen. Other elements including sulphur, nitrogen, and phosphorous may also be present in metabolites, but are usually limited to only a few atoms [24, 25]. As a consequence, metabolites are manifold in terms of chemical structure and properties [26]. For example, it is believed that the human metabolome consists of at least 3,000 metabolites [27], while the plant kingdom is estimated to have hundreds of thousands or more different metabolites [28].
Albrecht Kossel suggested stratifying metabolites into two classes: primary and secondary [29]. Primary metabolites are part of essential pathways, or are mandatory for a biological system to survive and reproduce (e.g. metabolites of the citric acid cycle or amino acids). Secondary metabolites allow a biological system to perform many more and different actions than it could with only its primary metabolites (e.g. pathogenicity or anti-bacterial functions) [30]. Thus,

secondary metabolites help organisms, in particular microorganism, to adapt to unconventional ecological habitats.

In metabolomics research two fundamental concepts are distinguished: targeted and untargeted approaches. Targeted approaches investigate only a small number of known metabolites and require either reference standards or authentic MS/MS data of the investigated metabolites [26; termed metabolite target analysis]. Determination of absolute metabolite concentration, however, in general requires the respective metabolites to be available as authentic reference standards. On the contrary, untargeted metabolomics approaches are aimed at capturing all metabolites present in a biological sample and also utilise information of unknown metabolites that usually constitute the vast majority of all detected compounds [31]. Consequently, only untargeted metabolomics applications can probe the entire metabolic space of the biological system under investigation at once. As with untargeted metabolomics approaches, the majority of all detected metabolites cannot be identified since only few reference standards are available. As a result, only relative metabolite concentrations can be determined and used to compare the different experimental groups in untargeted metabolomics approaches.
An obstacle in metabolomics approaches is the large chemical diversity of metabolites (e.g. fatty acids, sugars, polar and apolar metabolites). This makes it impossible to analyse an entire metabolome with just a single analytical protocol or platform. Thus, complementary techniques (e.g. gas chromatography coupled to mass spectrometry (GCMS), liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS), nuclear magnetic resonance (NMR)) have to be utilised conjointly and all data needs to be integrated for a more holistic annotation of a respective metabolome [32].
LC-HRMS is the most commonly utilised analytical technique for the study of secondary metabolites in untargeted metabolomics research. It is highly customisable (e.g. different chromatographic separation, eluents, and ionisation methods) making it fit for nearly any purpose [33]. Furthermore, most LC-HRMS systems (e.g. Orbitrap, Time-of-Flight) are sophisticated and allow a high reproducibility on the same instrument given appropriate analytical protocols and carefulness from experienced and trained experts [26]. The analytical protocols, workflows, and the LC-HRMS platform are generally adapted and optimised to allow for the simultaneous detection of as many metabolites as possible present in a single sample.

Typically, untargeted metabolomics experiments are performed with the aim of investigating the metabolic composition of an altered, biological system. The perturbations of the systems under investigation can be manifold (e.g. genetic mutations, environmental stress, nutrition depletion). Subsequently, the differing samples are compared to the same biological system that is not exposed to the studied factors and therefore serves as a control. Comparative quantification of the affected and control samples and statistical analysis are then performed to select those metabolites that are highly likely to be affected or responsible for these differences.

1.1. Conventional data processing strategies

Untargeted metabolomics research performed with modern LC-HRMS instrumentation produces huge amounts of raw-data. Depending on the type of instrument used for LC-HRMS analyses, a simple experiment comprising of two different conditions with five biological replicates each (n_{total}=10 samples) can range from a few hundred megabytes to some gigabytes, even if roughly comparable instrument settings are used (e.g. same chromatography and mass-to-charge ratio (m/z) range). In general, more powerful instruments that scan faster, have higher mass accuracy, and are more sensitive mainly caused this increase, which is likely to continue in the near future. While for targeted approaches semi-manual verification, adjustment, and analysis of the metabolites of interest are feasible, this certainly cannot be done in an untargeted fashion aimed at extracting as many metabolites as possible. Therefore, sophisticated, automatic data processing tools that detect and process all metabolite-derived signals and chromatographic peaks are necessary. Popular software for this task are XCMS [34-36] in combination with CAMERA [37], MzMine [38, 39], OpenMS [40, 41], MetAlign [42, 43], Lipid Data Analyzer [LDA; 44], Maven [45], and mzMatch [46]. These tools analyse the raw-data and holistically report all detected metabolites in the LC-HRMS data.

For example, XCMS is the most commonly used tool for automated data evaluation for untargeted metabolomics research. In a first step the tool clusters all m/z values detected in the entire LC-HRMS data into so called Regions of Interest (ROIs). For each such ROI, which is within a certain m/z error range and thus includes different isomers of metabolites with the same chemical formula, chromatographic peaks representing different isomers are then searched for. Each such detected chromatographic peak is reported to the user as a feature representing a metabolite-derived ion consisting of a certain m/z, retention time, and intensity value. Subsequently, CAMERA [37] annotates the relationship between different features originating from the same metabolite. To this end, associated features are convoluted using a) the m/z difference between two features originating from different 1) isotopologs, 2) ion species, or 3) in-source fragments of the same metabolite, and b) the chromatographic peak shapes of two coeluting features that are highly similar if they originate from the same metabolite. The result of a CAMERA analysis is a comprehensive list of all detected features most likely belonging to the same metabolites. Furthermore, using the detected isotopic patterns the chemical formulas of the observed ions can be predicted to a certain extend.

Detecting all metabolite-derived signals in complex LC-HRMS data has become a routine task using the before mentioned software solutions. However, these tools cannot account for certain limitations of the LC-HRMS data:

i. **Contaminants and substances of non-biological origin**
   LC-HRMS analysis cannot distinguish between substances of biological origin (i.e. metabolites) and such of non-biological origin (e.g. contaminants or background signals). For example, typical contaminants, which may originate from the analyti-

cal process, are plasticizer from the used laboratory equipment [47].

ii. **Matrix effects**
Different ion suppression and enhancement effects (matrix effects) may distort relative metabolite abundances across all analysed experimental conditions and are a potential source of error during subsequent statistical analysis [27, 48-50].

iii. **Metabolite annotation**
Annotation of unknown metabolites usually starts from accurately measured m/z values and the prediction of molecular formulas of the metabolites under investigation. To determine the number of atoms of a specific element, the relative abundances of different isotopologs need to be accurately quantified by the used LC-HRMS instrument and data processing software [51]. A deviation of as little as 1% already increases or decreases the number of carbon atoms by 1, however, typical LC-HRMS instruments have higher uncertainties [52].

In general, untargeted metabolomics experiments are aimed at finding metabolites that are characteristic for certain conditions (e.g. different genotypes, metabolic stress due to an environmental impact, a disease or infection). Thus, after feature detection, convolution, and relative quantification, statistical analysis with the aim of selecting those metabolites affected or responsible for these differences is performed [53, 54]. As an initial check principal component analysis (PCA) and hierarchical cluster analysis (HCA) are often utilised as a first check to investigate, whether the analysed samples separate into the expected experimental conditions followed by more in-depth multi- and univariate statistical investigations. Comprehensive literature on data pre-processing and statistical investigation in untargeted metabolomics research is available, including for example Liland [53], Issaq [54], Kruve [55], Broadhurst [56], Goodacre [57], Madsen [58], van den Berg [59], and Wehrens [60].

1.2. Stable isotopic labelling

Stable isotopic labelling (SIL) is the process of artificially creating molecules, which predominantly are enriched with a stable isotope that is not the most abundant one in nature (e.g. ^{13}C or ^{15}N) [1]. SIL has already been used in life science related applications before it has been utilised for untargeted metabolomics research [61-65]. Moreover, it has been and is intensively used in fluxomics applications where the aim is to gain insight into the rates of metabolic transformation of different metabolites [66, 67]. However, most commonly fluxomics applications are targeted approaches and only investigate certain pathways of a biological system and are not addressed by the presented work at all. Here, labelled metabolites or labelled parts of metabolites need to have a high degree of isotopic enrichment (≥ 97%) with the respective isotope used for labelling.
Native and ^{13}C-, ^{15}N- or ^{34}S-labelled metabolites possess very similar physico-chemical proper-

ties. The minor differences, however, are not of relevance in the presented context. Thus, native and ^{13}C-, ^{15}N- or ^{34}S-labelled metabolites are not separated during LC and show perfect coelution [68]. Since ions of native and labelled metabolites have a slightly different mass, their different isotopologs are separated in the mass spectrometer (e.g. native phenylalanine with a monoisotopic mass of 165.078724 u and globally ^{13}C-labelled phenylalanine with a mass of 174.108874 u). As a consequence, mixtures of native and ^{13}C-labelled metabolites elute at the same retention time but show distinct isotope patterns in LC-HRMS data. The observed m/z difference between the native and labelled metabolite-derived ions allows calculating the number of atoms of the labelling-isotope without the need to approximate it from different isotopologs. Additionally, in combination with a sophisticated sample-pooling scheme before LC-HRMS analysis, experiment- and metabolome-wide internal standardisation can be achieved. This efficiently accounts for different matrix effects observed for the studied samples or MS detector sensitivity drifts. Therefore, this technique is perfectly suited for the study of the entire via LC-HRMS accessible metabolome of various biological systems (full metabolome labelling) or the study of the metabolic fate of certain tracer substances (tracer-fate studies) (publication #1)

For example, Giavalisco and colleagues used SIL in untargeted metabolomics experiments involving *Arabidopsis thaliana* plants [69, 70]. The plants were grown in a hermetically sealed chamber, where the environment had been enriched with ^{13}C, ^{15}N, and ^{34}S. Consequently, all the plant's metabolites predominantly consisted of the respective isotopes. Together with *A. thaliana* extracts that had been cultivated under native, label-free conditions LC-HRMS measurements showed distinct isotope patterns for all metabolites. These patterns allowed determining the number of carbon, nitrogen and sulphur atoms in each metabolite and subsequently improved the annotation of the unknown metabolites with chemical formulas. More recently, Cano and colleagues used ^{13}C- and ^{15}N-labelling for the unambiguous detection and annotation of twenty-one secondary metabolites of *Aspergillus fumigatus* [71]. In respect of relative metabolite quantification, Giavalisco and colleagues showed that SIL improved the comparability of different samples by accounting for unequal matrix effects [72]. Furthermore, tracer-fate studies aiming at determining the metabolic fate of studied tracer substances already utilised SIL in the late 1980s [73, 74]. Native and ^{13}C-labelled tracer substances are almost equally metabolised by the same biotransformations and thus all biotransformation products contain either the native or the labelled tracer [75]. For example, Anderson [76] used deuterium-labelled steroids in pregnant women and Li [77] used deuterium-labelling in combination with a PCA guided approach to detect known and unknown biotransformation products of the drug Tempol.

Currently, only few software tools are available that are designed for processing of SIL-derived metabolomics data (e.g. NTFD [78, 79], X^{13}CMS [80], and mzMatch-Iso.R [81]).
NTFD (Non-targeted Tracer Fate Detection) is a software tool designed for the study of the metabolic fate and flux of primary metabolites (e.g. sugars) with GCMS and currently does not support LC-HRMS data of secondary metabolites. In contrast to many other fluxomics applications, it does not require a list of afore known metabolites but rather detects tracer-derived

metabolites in an untargeted manner.
The latter two tools, X^{13}CMS and mzMatch-Iso.R, are capable of processing SIL-derived LC-HRMS data. They are designed to comprehensively annotate the metabolic space. Subsequently, they select those metabolites that show significant differences in the isotope patterns of the biological conditions under investigation. Thus, these tools require separate LC-HRMS analysis of samples consisting of native and labelled metabolites.

In their current forms, X^{13}CMS and mzMatch-Iso.R do not support the comprehensive detection of all metabolites that are present as native and highly isotope-enriched metabolites only. Furthermore, the experimental setup necessary for these tools does not allow experiment- and metabolome-wide internal standardisation nor does it allow deriving the total number of labelling-isotopes present in each detected metabolite or biotransformation product. Nevertheless, these tools are valuable for the investigation of partly labelled metabolites as well as the elucidation of putative novel pathways and their metabolic fluxes.

1.3. Aims of this thesis

Xiaojing Huang and colleagues, who just recently presented the X^{13}CMS workflow for SIL assisted untargeted metabolomics, write: "*Studies of isotopically labelled compounds have been fundamental to understanding metabolic pathways and fluxes. They have traditionally, however, been used in conjunction with targeted analyses that identify and quantify a limited number of labelled downstream metabolites*" [80]. This emphasises the need for SIL assisted untargeted metabolomics workflows. The limiting factor in this respect is the availability of software tools for the automated analysis of the acquired LC-HRMS data.

At the beginning of the author's doctoral study no software tools for the automatic processing of LC-HRMS derived metabolomics data from SIL-assisted experiments were available. In view of the afore described challenges and limitations, the rationales of the presented work was the development and implementation of novel tools that enable the automated data processing of SIL assisted and LC-HRMS based metabolomics data in a) full metabolome labelling experiments and b) tracer-fate studies. More specifically, the software should provide the following functionalities:

i. The holistic and untargeted detection of all analytical signals of truly biological origin (i.e. native and isotopically labelled metabolite-derived ions)
ii. The holistic and untargeted detection of all biotransformation products derived from labelled endogenous or exogenous tracer compounds, which are transformed by secondary metabolic processes into other metabolites
iii. Determine the total number of atoms of the labelling-isotope in each feature pair
iv. Track the same metabolites over all experimental samples
v. Calculate internal standardisation corrected relative metabolite abundances
vi. Group different ion species derived from the same metabolite

vii. Support different stable isotopes used for labelling (e.g. ^{13}C, ^{15}N, ^{34}S)

viii. Process data from different LC-HRMS platforms (e.g. Orbitrap or Time-of-flight instruments)

ix. Handle positive, negative, and fast polarity switching ionisation mode LC-HRMS data

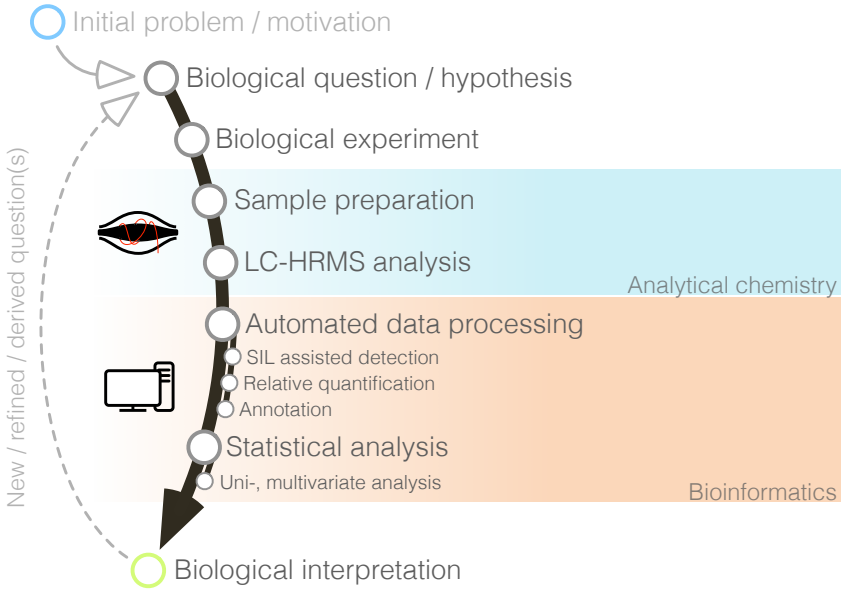


Figure 1 | Illustration of a typical workflow for untargeted metabolomics experiments (black arrow) that served as the basis for the developed workflow(s). The presented doctoral thesis comprises of the implementation of a software tool for automated processing of SIL-derived LC-HRMS data (Bioinformatics), while the analytical aspects of the workflows are detailed in the doctoral thesis of Bernhard Kluger.

These rationales were carried out in the context of the subproject SFB-3706 "Metabolomics of plant-Fusarium-interactions" in the research group "Metabolomics and Bioactive Compounds" that is headed by Dr. Rainer Schuhmacher (Center for Analytical Chemistry, University of Natural Resources and Life Sciences, Vienna) in collaboration with Dr. Gerhard Thallinger (Institute for Knowledge Discovery, Graz University of Technology). Together with the PhD student Bernhard Kluger, who is an analytical chemist, and other group members, two complete metabolomics workflow were developed and applied. These workflows include SIL assisted biological experiments, adapted sample preparation, LC-HRMS analysis, automated data processing, and statistical investigation (Figure 1). Both the analytical and bioinformatics workflows greatly benefitted from the parallel development and mutual inspiration. The workflows and the software tool were rigorously tested and utilised to study the entire LC-HRMS accessible metabolomes of wheat plants and filamentous fungi as well as the metabolic fate of different tracer substances in plants and cell suspension cultures.

This doctoral thesis focuses on the implemented software MetExtract while the detailed analytical workflows are described in the doctoral thesis of Bernhard Kluger.

2. Stable isotopic labelling of biological material

The process of SIL as used throughout this work is explained where the focus is placed on the developed workflows for SIL assisted full metabolome labelling experiments and tracer-fate studies as well as the formal definition of the unique characteristics in such experiments. Other SIL assisted approaches in untargeted metabolomics include for example derivatization/dansylation with isotopically labelled agents [82], stable isotope dilution assays [83, 84], or SIL assisted annotation of LC-HRMS /MS spectra [85]. These concepts are, however, beyond the rationales of the presented work.

For most elements one isotope (e.g. ^{1}H, ^{12}C, ^{32}S) is predominantly present in nature. It is called the principal isotope of the element. Some elements may have other stable isotopes with a different number of neutrons. Usually, these isotopes are present only in minor quantity (e.g. ^{2}H, ^{13}C, ^{34}S; see http://www.sisweb.com/referenc/source/exactmaa.htm; last accessed 17th September 2014).

In the presented work, SIL refers to the process of artificially creating molecules enriched with an isotope other than the principal one of the respective element [86]. For example, in nature ^{12}C is the principal isotope of carbon with a relative abundance of 98.9%, while the second most abundant stable isotope of carbon, which is ^{13}C, makes up only 1.1%. Artificial labelling with the isotope ^{13}C creates molecules that predominantly consist of ^{13}C atoms. Such isotopologs are almost impossible to find in natural environments. In general, all atoms of a particular element in a certain substance have the same probability to be replaced during the isotopic labelling process resulting in globally labelled molecules. On the contrary, in certain metabolites (e.g. biotransformation products of tracer substance under investigation) not all atoms of a particular element are replaced with the isotope used for labelling, but only a constant subpart or certain atom positions. Such molecules are partially labelled only since just certain atoms of the element, which is facilitated for labelling, can be labelled. Consequently, these molecules consist of parts having properties of either non-labelled, native or globally labelled substances. The degree of enrichment with the isotope used for labelling in this work is presumed to be very high (usually above 97%).

Metabolites labelled with ^{13}C, ^{15}N, or ^{34}S isotopes are not separated during LC from their native analogues since they posses nearly the same physico-chemical properties and thus show perfect coelution in the presented context [68, 87]. Yet, ions derived from labelled metabolites are separated in the mass analyser as they have different m/z values compared to their non-labelled, native analogues. Isotopes of hydrogen (^{2}H) and oxygen (^{18}O), which would also be perfectly suited for labelling, are not commonly used, since these two elements are known to cause problems during LC-HRMS analysis [68]. For example, deuterium-labelled molecules are known to chromatographically separate from their non-labelled pendants. Moreover, deuteri-

um and ^{18}O atoms can be exchanged with H and O atoms from surrounding, non-labelled water, which would reduce the labelling degree and result in only partially labelled molecules and blurred isotope patterns in the LC-HRMS data. Consequently, in the presented work, ^{13}C is the primary isotope used for labelling. Figure 2 illustrates examples of native and globally or partly ^{13}C-labelled molecules.

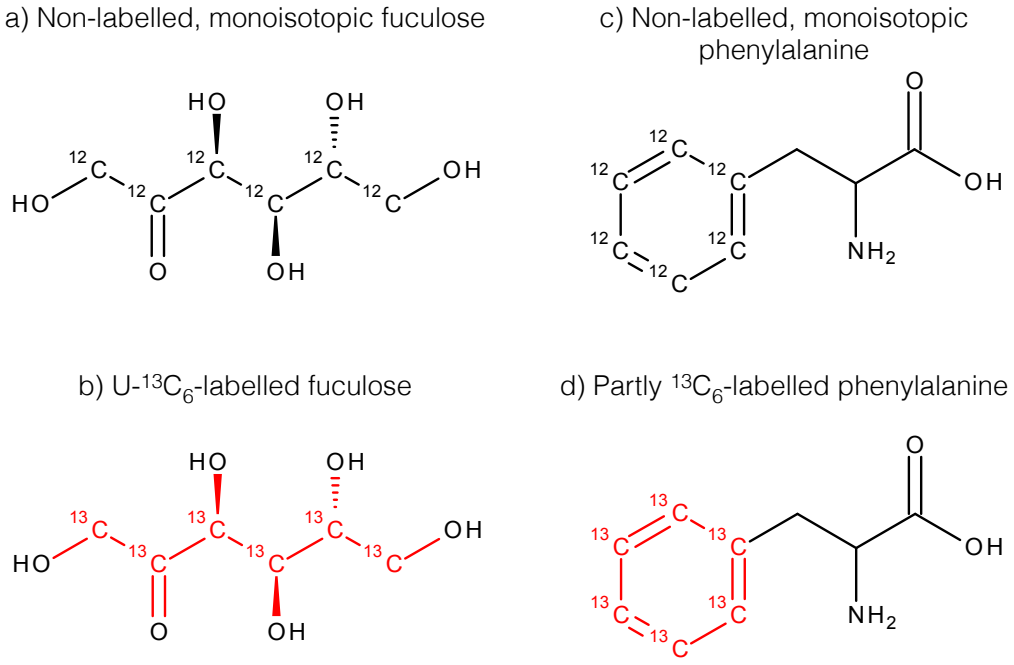


Figure 2 | Illustration of non-labelled, native (a and c) and globally (b) or partly (d) ^{13}C-labelled metabolite molecules. While the non-labelled fuculose molecule (a) contains only ^{12}C isotopes for all its carbon atoms, the globally ^{13}C-labelled fuculose molecule (b) contains only ^{13}C isotopes instead of ^{12}C. The second example depicts the aromatic amino acid phenylalanine. The native phenylalanine molecule (c) is made solely of ^{12}C isotopes, while only the aromatic ring of the phenylalanine molecule is $^{13}C_6$-labelled (d), thus this molecule contains ^{12}C as well as ^{13}C isotopes and is only partly labelled. Different isotopologs of the native or labelled molecules are not shown.

2.1. Full metabolome labelling – biological workflow

For the SIL assisted study of the entire metabolome of a respective biological system (e.g. fungi or wheat organs) it is necessary to enrich all its metabolites with a respective stable isotope. If such a setup can be achieved (e.g. cultivation of *F. graminearum* in parallel on native and globally ^{13}C-labelled glucose; publication #3), all metabolites of the studied biological system are present as native and globally labelled molecules. However, in many metabolomics applications such a co-cultivation is not possible. For example, globally ^{13}C-labelled wheat or maize plants could not be grown in-house since the required facilities were not available (publication #3). Thus, labelled material of respective wheat and maize plants had to be acquired commercially.

After the biological experiment has been performed, sample preparation is carried out separately for the native and labelled samples with the aim to extract the metabolites from the biological material for LC-HRMS analysis (e.g. protocol of De Vos [88]). Subsequently, all extracts from the labelled samples are pooled to gain a unified and globally labelled sample that contains all labelled metabolites from the different, studied samples. Alternatively, if the labelling cannot be achieved in parallel, a surrogate in form of labelled reference material may be used. An equal amount of the labelled, pooled sample or the labelled reference sample is then added to each of the samples containing the individual, native metabolite molecules of the non-labelled samples. After this mixing step, the samples contain native as well as globally labelled metabolite molecules. Since the labelled metabolites are present in equal amount in each sample, they can be used for metabolome- and experiment-wide internal standardisation by the developed software tools to account for different matrix effects during LC-HRMS analysis. These sample preparation steps are illustrated in Figure 3.

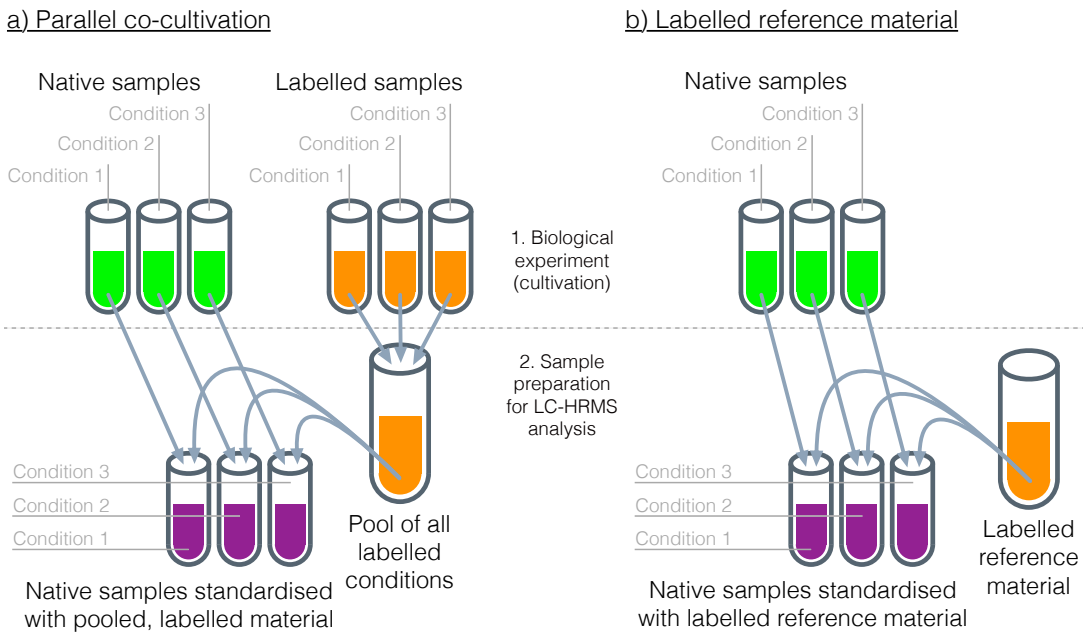


Figure 3 | Biological experiment (1.) and analytical sample preparation (2.) steps full metabolome labelling experiments. When parallel cultivation of the respective biological system under investigation can be achieved, each experimental condition is cultivated under native environmental conditions (green samples) as well as such enriched with the isotope used for labelling (orange samples; e.g. co-cultivation on native and labelled glucose) (a). If such a parallel co-cultivation cannot be achieved (e.g. lack of required facilities or the cultivation substrates), a labelled reference material has to be commercially acquired (b). After metabolite extraction, an equal amount from either the labelled, pooled sample or the labelled reference material is added to each individual, native sample. Replicates of each condition are not shown.

2.2. Tracer-fate studies – biological workflow

In the presented work examining the metabolic fate of a tracer substance in a biological system under investigation is carried out with native and labelled tracers. Both forms are metabolised by identical biotransformation reactions. Consequently, all metabolites derived from the studied tracer are present in their native as well as a partly labelled form, while any other metabolites of the studied, biological system that are not derived from the tracer are only present as native metabolites. This allows easily discriminating between metabolites derived form the tracer substance under investigation and such that are not. Thus, both the native and labelled forms of the tracer need to be available to the biological system. In this respect two types of tracer substances are distinguished:

i. **Exogenous tracer**
   An exogenous tracer is a substance that is typically not present in the investigated biological system but only in certain situations (e.g. toxins or drugs). Thus, to study such a tracer, the substance must be supplied as both the native and labelled tracer. Figure 4a illustrates the biological workflow for an exogenous tracer.

ii. **Endogenous tracer**
   On the contrary, an endogenous tracer is a substance typically present in the biological system under investigation. Thus, native molecules of the tracer are already present and only the labelled tracer has to be supplied during the biological experiment. Figure 4b illustrates the biological workflow for an endogenous tracer.

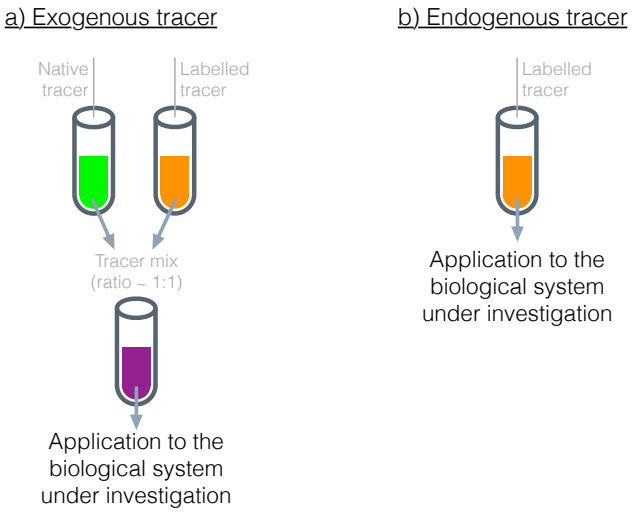


Figure 4 | Biological workflow for the study of an exogenous (a) or an endogenous tracer (b) substance. While an exogenous tracer must be supplied to the biological system as a mixture of both native and labelled forms, an endogenous tracer only needs to be supplied as a labelled form.

2.3. Nomenclature for isotope patterns of native and labelled compounds

Isotope patterns in LC-HRMS data originate from different isotopologs of the same metabolite ions. Different isotopologs represent the same substance, but are different in their isotopic constitution. Consequently, all isotopologs of a substance have the same elemental but a different isotopic composition and thus a different mass.

In this work, metabolites are labelled with an isotope of a single element only. The isotope used for artificially creating the labelled molecules is termed the labelling-isotope. Consequently, the element of the used labelling-isotope is named the labelling-element. For example, in a SIL experiment, which uses ^{13}C for labelling, the labelling-element is carbon while the labelling-isotope is ^{13}C.

The isotopolog of a substance, which contains only the principal isotopes of all its elements, is termed the monoisotopic isotopolog. For example, the monoisotopic isotopolog of the substance fuculose (Figure 2a) has the sum formula $^{12}C_6{}^{1}H_{12}{}^{16}O_5$.
Labelled metabolites have a pendant similar to the monoisotopic isotopolog, which contains only the principal isotopes of all those elements not utilised for labelling. Additionally, depending on the performed metabolomics experiment the following restrictions also apply to the respective isotopolog:

i. **Full metabolome labelling experiments**
   All atoms of the respective labelling-element are replaced with the labelling-isotope. In full metabolome labelling applications this isotopolog is termed the uniformly labelled isotopolog of a globally labelled metabolite. Thus, for each element the uniformly labelled isotopolog of a globally labelled metabolite consists of only a single isotope for each of its elements. For example, the uniformly ^{13}C-labelled isotopolog of fuculose (Figure 2b) has the sum formula $^{13}C_6{}^{1}H_{12}{}^{16}O_5$.

ii. **Tracer-fate studies**
   In tracer-fate studies biotransformation products derived from a labelled tracer represent a mixed isotopolog form. Such molecules have atomic positions, which are labelled and others, which are not. In tracer-fate studies this isotopolog is termed the consistently labelled isotopolog of a biotransformation product. Such an isotopolog contains only the labelling-isotope for all those atoms that can be labelled (i.e. originate from the tracer) but the principal isotope of the labelling-element for all those atoms of the labelling-element that cannot be labelled. For example, the phenylalanine molecule in Figure 2d is partly labelled only. Its aromatic ring is globally ^{13}C-labelled. Thus, the consistently labelled isotopolog of this

labelled phenylalanine molecule has the sum formula $^{13}C_6{}^{12}C_3{}^{1}H_{11}{}^{16}O_2{}^{14}N$ and all 6 carbon atoms of the aromatic ring are labelled.

Consequently, the monoisotopic and the uniformly or consistently labelled isotopolog of a metabolite only differ in the number of isotopes of the facilitated labelling-element.

LC-HRMS analysis of native and ^{13}C-, ^{15}N-, or ^{34}S-labelled metabolite molecules does not separate the different metabolite isotopologs during liquid chromatography. Thus, the native and labelled metabolites elute with the same retention time. During chromatographic separation, the eluent is constantly directed into the ionisation device (e.g. electrospray-ionisation (ESI) used throughout this thesis). There the metabolite molecules are ionised and the same ion species are created for both the native and labelled metabolites (e.g. protonated ions $[M+H]^+$, deprotonated ions $[M-H]^-$, sodium adducts $[M+Na]^+$). Subsequently, in the mass spectrometer the native and labelled metabolite-derived ions are separated, since their mass-to-charge ratio (m/z) values are different. This m/z difference corresponds to the number of labelling-isotopes in the respective metabolite ions.
Metabolite-derived ions in the presented workflows for SIL experiments show at least two distinct isotope patterns in LC-HRMS data originating from different isotopologs of native and labelled metabolite ions respectively. In the presented work, the labelling was predominantly performed with ^{13}C, which results in very distinct isotope patterns compared to labelling with other isotopes (e.g. ^{15}N or ^{34}S). These ^{13}C-labelling-isotope patterns are defined as follows:

- M corresponds to a monoisotopic, native metabolite-derived ion. Usually, the monoisotopic isotopologs represent the most abundant signals of the respective isotope patterns for native metabolite ions.

- M' denotes different isotopologs depending on the performed SIL experiment; in full metabolome labelling experiments M' refers to the uniformly, ^{13}C-labelled isotopolog of a globally labelled metabolite, while in tracer-fate studies M' corresponds to the consistently ^{13}C-labelled isotopolog of a partly labelled biotransformation product. This nomenclature is based on the fact that M' is also usually the most abundant signal among its respective isotope pattern of the ^{13}C-labelled metabolite.

- M+x denotes the x^{th} carbon-isotopolog of native metabolite-derived ions. x ^{12}C isotopes of the monoisotopic isotopolog are replaced with x ^{13}C isotopes. However, this replacement happens also by chance as approximately 1.1% of all carbon atoms in nature are ^{13}C and is not introduced by the SIL process.

- M'-x denotes for the x^{th} carbon-isotopolog of the labelled metabolite-derived ion M'.

- In full metabolome labelling experiments involving globally labelled metabolites, M'-x corresponds to that labelled isotopolog ion, in which x ^{13}C isotopes of all atoms of the labelling-element are replaced with x ^{12}C isotopes.
- In a biotransformation product, M'-x corresponds to that isotopolog ion, in which x ^{13}C isotopes of a consistently labelled biotransformation product are replaced by x ^{12}C isotopes. These exchanges only happen at carbon atom positions, which can be labelled according to the used labelled tracer.

- M'+x denotes the x^{th} carbon isotopolog of the non-labelled part of a partly labelled biotransformation product. x ^{12}C isotopes are replaced by x ^{13}C isotopes. As a consequence, M'+x can only be present for such biotransformation products, which are a conjugation of a labelled tracer with a native moiety originating from the biological system under investigation. This replacement happens by chance as approximately 1.1% of all carbon atoms in nature are ^{13}C and is not caused by the SIL process and occurs only at carbon atom positions, which cannot be labelled according to the used labelled tracer.

- M+x/M denotes the observed ratio of the relative abundances of the isotopologs M+x and M. It originates from the fact that approximately 1.1% of all carbon atoms in natural environments are ^{13}C isotopes.

- M'-x/M' denotes the observed ratio of the relative abundances of the isotopologs M'-x and M'. It originates from the fact that the labelling-isotope has a certain change to incorporate ^{12}C isotopes instead of ^{13}C since the source of labelling-isotopes are typically not enriched entirely with the labelling-isotope (e.g. only 98.5%).

- M'+x/M' denotes the observed ratio of the relative abundances of the isotopologs M'+x and M'. Like the ratio M+x/M it does not originate from the labelling process but from the fact, that approximately 1.1% of all carbon atoms in nature are ^{13}C isotopes.

- M:M' denotes the observed ratio of the relative abundances of the isotopologs M and M'.

- A carbon-isotope pattern is depicted in the form |A, B, ...|. It starts from the isotopolog A and continues to B. For example the ^{13}C-isotope pattern of a native metabolite ion is denoted with |M, M+1, ...| while the complementary ^{12}C-isotope pattern of its uniformly ^{13}C-labelled counterpart is denoted with |M', M'-1, ...|.

2.4. Characteristics of native and SIL-derived LC-HRMS data

For the implementation of a software capable of detecting metabolite present as native and labelled forms, the respective isotope pattern in the acquired LC-HRMS data need to be formally defined. In the following these patterns are elucidated.

SIL can be performed with any element consisting of at least two stable isotopes. The most suited element for labelling is carbon since it is present in virtually every metabolite and most have a high number of carbon atoms. Other elements like sulphur or nitrogen may also be used, but only a minor number of all metabolites contain these elements [24]. Figure 5 shows the difference on a MS scan level between ^{13}C- and ^{15}N-labelling for the same ion species of a native and globally labelled metabolite.

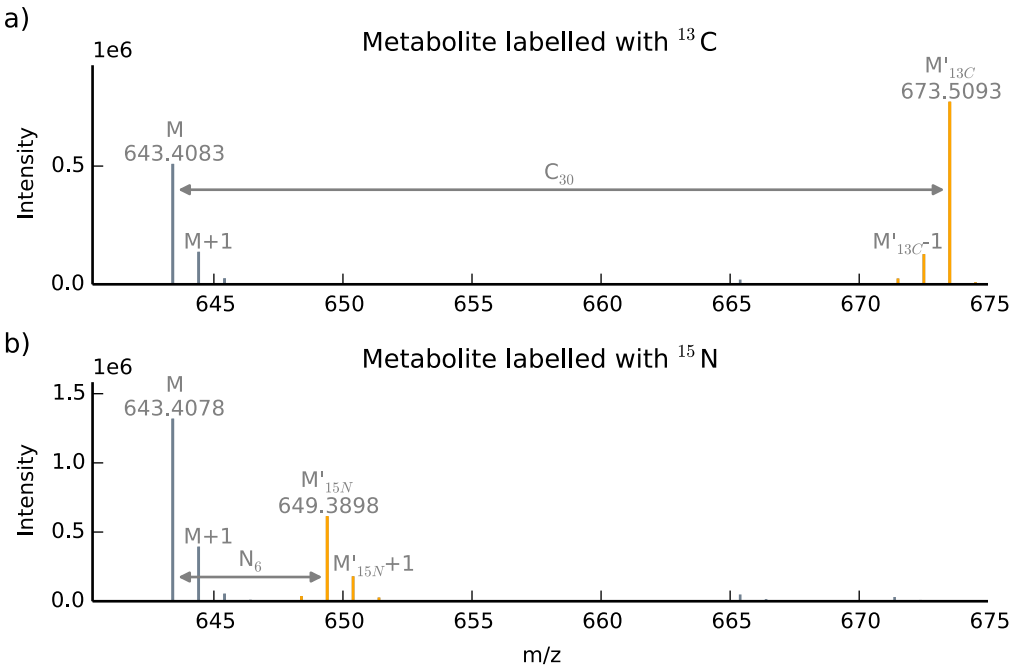


Figure 5 | In a) the mass spectrum shows the different isotopologs of an ion of a native (grey) and a globally ^{13}C-labelled (orange) metabolite. From the *m/z* difference between the native, monoisotopic (M; *m/z* 643.4083) and the uniformly ^{13}C-labelled (M'$_{13C}$; *m/z* 673.5093) ions the total number of carbon atoms in that ion is determined to be 30. In b) the labelling of the respective metabolite is performed with ^{15}N instead of ^{13}C. Consequently, the *m/z* difference between the native, monoisotopic (M; *m/z* 643.4078) and the uniformly ^{15}N-labelled (M'$_{15N}$; *m/z* 649.3898) ions corresponds to 6 nitrogen atoms. As a result of this separately performed labelling, this unknown metabolite-derived ion can be annotated with C_{30} and N_6 and a charge of 1.

2.4.1. Carbon-13 labelling

In ^{13}C-labelling experiments the observed carbon isotope pattern of a metabolite is a composition of two individual carbon isotope patterns originating from either the native and the ^{13}C-labelled metabolite ions. An example is depicted in Figure 5a. In general, the isotopolog patterns of native and highly isotopic enriched metabolites do not overlap in the presented work, but would, if, for example, the number of carbon atoms in the metabolite molecule is low (e.g. less than 3 carbon atoms) and/or if the degree of enrichment of ^{13}C in the labelled form is not sufficient (e.g. an enrichment of less than 10%). The ions of the monoisotopic, native isotopolog M (m/z 643.4081) and the uniformly ^{13}C-labelled isotopolog M' (m/z 673.5090), which are usually the most abundant signals among their respective isotope patterns, allow deriving the total number of carbon atoms for a respective metabolic feature since the m/z difference between M' and M is proportional to the number of carbon atoms and the charge number of the respective ion. The carbon-isotope pattern of the native metabolite-derived ions |M, M+1, ...| starts at the monoisotopic isotopolog M and descends towards higher m/z values, which is a result of the replacement of ^{12}C isotopes with naturally occurring ^{13}C isotopes. On the contrary, since in a globally ^{13}C-labelled metabolite all carbon atoms are of the heavier isotope ^{13}C, the carbon-isotope pattern |M', M'-1, ...| of an ion derived from a labelled metabolite descends towards lower m/z values.

The relative abundances of a carbon-isotope pattern of a metabolite can be calculated with Equation 1. It can be used to determine the pattern of a native or a uniformly/consistently ^{13}C-labelled metabolite. a denotes the total number of carbon atoms in the respective metabolite, while s stands for the xth carbon isotopolog in M+x or M'-x and even M'+x. p denotes the degree of isotopic enrichment with the more abundant isotope (^{12}C in native metabolites and ^{13}C in ^{13}C-labelled metabolites respectively). In case of a native metabolite, p has the value 0.9893 as this is the relative abundance of ^{12}C isotopes in nature. The second stable isotope of carbon is not explicitly specified since typically those two isotopes make up nearly 100% of all carbon atoms and the remaining, non-stable carbon-isotopes can be neglected for the presented purpose. Consequently, P(6, 1, 0.9893), which evaluates to 6.45%, gives the abundance of the first carbon isotopolog M+1 of a native fuculose isotopolog with the sum formula $^{13}C^{12}C_5{}^{1}H_{12}{}^{16}O_5$ relative to its monoisotopic isotopolog. On the contrary, the degree of isotopic enrichment in ^{13}C-labelled metabolites is not constant and varies between different suppliers of isotopic enriched material. For example, P(6, 1, 0.995), which evaluates to 3.02%, gives the abundance of the first carbon isotopolog M'-1 of a globally ^{13}C-labelled fuculose isotopolog with the sum formula $^{13}C_5{}^{12}C^{1}H_{12}{}^{16}O_5$ and 99.5% degree of ^{13}C enrichment relative to its uniformly ^{13}C-labelled isotopolog.

$$P(a,s,p) = \frac{p^{a-s}(1-p)^{s}\binom{a}{s}}{p^{a}} \quad (1)$$

In biotransformation products, which consist of a labelled tracer and a native moiety, Equation 1 needs to be evaluated separately for its two parts. For example, in the partly labelled phenylalanine shown in Figure 2d P(6, 1, 0.993), which evaluates to 4.23%, calculates the abundance of M'-1 relative to M', while P(3, 1, 0.9893), which evaluates to 3.24%, calculates the abundance of M'+1 relative to M'. On the other hand, the abundance of M+1 relative to M of the native phenylalanine is calculates with P(9, 1, 0.9893), which is 9.73%.

With Equation 1 the probability of a non-labelled, native metabolite consisting solely of ^{13}C isotopes can be calculated. For example, isotopologs of DON ($C_{15}H_{20}O_6$) consisting only of principal ^{12}C isotope of carbon have a probability of 0.85 among all DON isotopologs. On the contrary, DON-isotopologs consisting only of ^{13}C isotopes have a probability of 2.8×10^{-30} to occur in natural environments. To express it differently, in 178 tons of pure DON only a single DON molecule would consist solely of 15 ^{13}C isotopes. Therefore, fully labelled metabolites are virtually not present in natural environments and thus can be neglected completely.

2.4.2. Nitrogen-15 or sulphur-34 labelling

Labelling with isotopes other than ^{13}C (e.g. ^{15}N or ^{34}S) also results in distinct isotope patterns. However, those are not as unique as the isotope pattern derived for ^{13}C-labelling experiments. For example, the abundance of the principal nitrogen isotope ^{14}N is 99.5% and most nitrogen-containing metabolites consist only of few nitrogen atoms (e.g. less than 10). As a result, such metabolites do not show a prominent nitrogen-isotope pattern in LC-HRMS data, which is additionally convoluted with the more much more prominent carbon-isotope patterns of the metabolites. An example of a ^{15}N-labelled metabolite ion is shown in Figure 5b. The mass peak at m/z 643.4081 denotes the native monoisotopic metabolite ion consisting solely of ^{14}N, while the mass peak at m/z 649.3899 refers to the uniformly ^{15}N-labelled metabolite ion. From this m/z difference the number of nitrogen atoms is calculated to be 6. The nitrogen-isotope pattern for both the native and the ^{15}N-labelled metabolite ions are therefore quite low abundant and not obvious. Yet, the isotope patterns of carbon are clearly present for both the native and the ^{15}N-labelled metabolite ions. Both carbon-isotope patterns descend towards higher m/z values since they are of natural origin in both the native and ^{15}N-labelled form and are not a result of the SIL process. Also, since the ^{15}N-labelling does not alter the carbon composition in the ^{15}N-labelled ions, both carbon-isotope pattern abundances are the same in terms of relative isotopolog ratios.

2.4.3. Full metabolome labelling

Full metabolome labelling applications are aimed at detecting and annotating all LC-HRMS accessible metabolites of a biological system under investigation. During sample preparation the native and the globally labelled biological samples are merged and successively measured by LC-HRMS. Consequently, each metabolite of the studied biological system shows distinct isotope patterns in the LC-HRMS data. Figure 6 shows MS spectra and extracted ion chromatograms (EICs) derived from three different ion species of the same metabolite in a full metabolome labelling experiment.

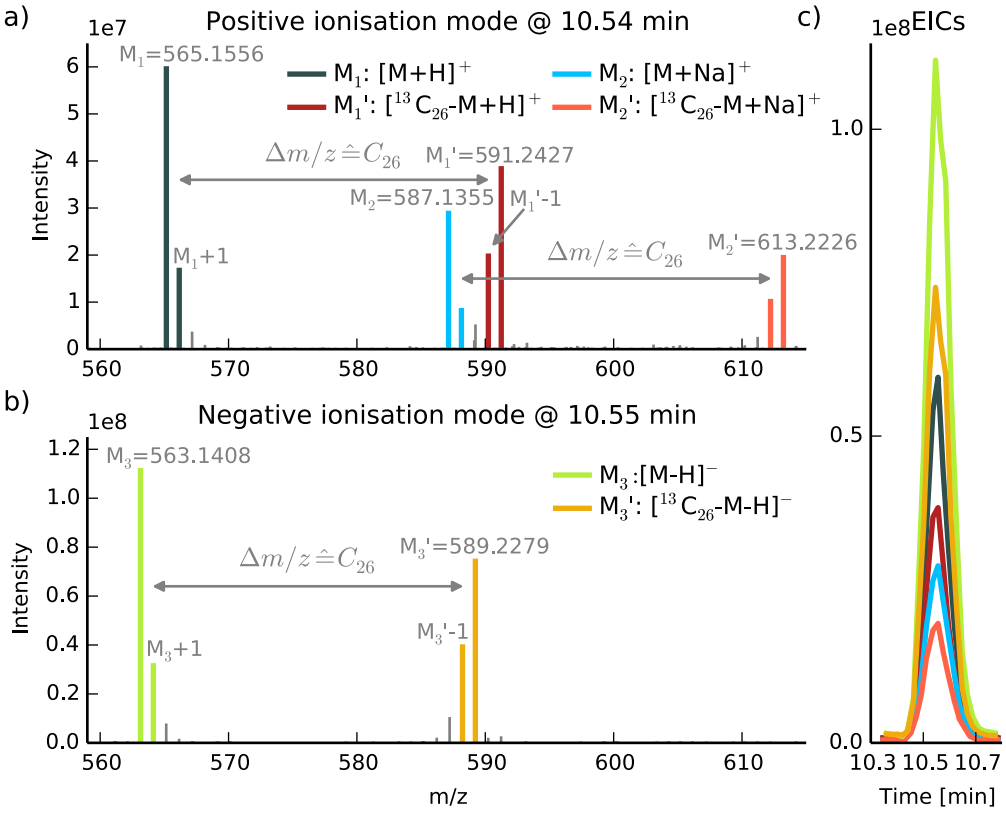


Figure 6 | Three feature pairs derived from the same unknown metabolite recorded in positive and negative ionization modes. a) shows a part of a positive ionization mode mass spectrum while b) shows a part of a negative ionization mode mass spectrum. In both MS spectra the unique isotope patterns |M_i, M_i+1, ...| and |M_i', M_i'-1, ...| for the three different ion species of the metabolite ions are clearly present. The mass difference from all three feature pairs allows annotation of the metabolite with 26 carbon atoms. Since the labelling was performed with ^{13}C, all isotopologs of the metabolite perfectly coelute, which is shown in c). Additionally, the ratios of M_i:M_i' are also very similar (approximately 1.5 for all three metabolite ions). From these three ions the monoisotopic mass of the uncharged metabolite was calculated to be 566.1480 u and it was annotated with 26 carbon atoms at a retention time of 10.55 minutes and a charge of 1.

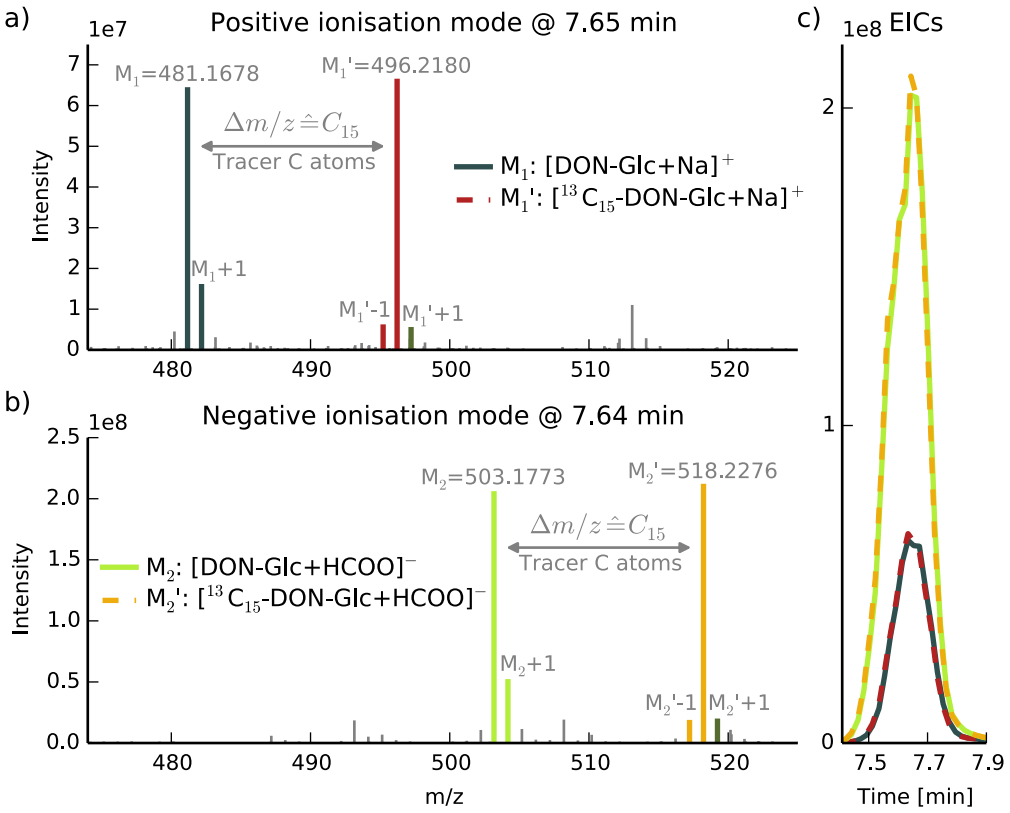


Figure 7 | Two different feature pairs of a biotransformation product are shown. a) and b) show two mass spectra of the same metabolite successively recorded in positive and negative ionisation modes. Both mass spectra show the distinct, carbon isotope patterns of the recorded biotransformation product ions. This metabolite contains the intact tracer molecule (DON), which is conjugated to a glucose molecule (Glc) originating from the biological system. The number of carbon atoms (15) calculated from the *m/z* difference between M_i' and M_i only represents the number of carbon atoms still present in the tracer-derived part of the respective biotransformation product, but it does not allow calculating the total number of carbon atoms in this biotransformation product. However, the presence of the mass peaks M_i'+1 indicate that the conjugated, native moiety contains several carbon atoms itself. The applied ratio of the pure native and the globally ^{13}C-labelled xenobiotic tracer substance (1:1/v:v) is well preserved in the LC-HRMS data. Furthermore, native and labelled biotransformation products also perfectly coelute as shown in c). The relative abundances of both forms were so similar in each recorded MS scan that the chromatographic peaks of the native and labelled forms perfectly overlay. Thus, for improved illustration, the EICs of the labelled biotransformation product ions are dashed.

2.4.4. Tracer-fate studies

Isotope patterns derived from native and labelled biotransformation products in tracer-fate studies show an important difference when compared to isotope patterns of native and globally labelled metabolites in full metabolome labelling approaches. While the tracer-derived part of such a biotransformation product is present as both a native and a labelled form, the moiety is always present in native form only. Therefore, in such biotransformation products, M' refers to a mixed isotopolog ion showing characteristics of native and labelled molecules, which af-

fects its entire isotope pattern. While the isotope pattern |M, M+1, ...| observed for the native metabolite reflects the total number of atoms of the labelling-element in this metabolite, the isotope pattern |M', M'-1, ...| does not. It only corresponds to the number of atoms of the used labelling-isotope present from the remaining tracer in this biotransformation product. Additionally, a third isotope pattern |M', M'+1, ...|, is clearly present for such biotransformation products that have additional atoms of the labelling-element in the conjugated moiety. Consequently, this isotope pattern only corresponds to the number of atoms of the used labelling-element present in the non-labelled parts of the biotransformation product (Figure 7). While the number of labelling-isotopes can be derived from the m/z difference between M' and M, the number of atoms of the used labelling-element present in the native moiety can only be estimated from the ratio M'+1/M', which is likely to be distorted depending on the used LC-HRMS platform.

For tracers that are exogenous to the studied biological system (e.g. toxins, drugs), the ratio M:M' of a putative biotransformation product will be approximately equal to the ratio observed for the applied, pure tracers, since the rates of biotransformation of native and labelled substances are nearly equal [2, 72]. On the contrary, if the applied tracer is an endogenous substance to the biological system under investigation (e.g. labelled amino acid), the observed ratio M:M' will usually be equal or higher than the ratio of the applied, pure labelled tracer since more of the native tracer is present.

3. Materials and methods

A novel software tool for the automated processing of SIL assisted and LC-HRMS derived metabolomics data was implemented. This software, which was named MetExtract, was implemented in the Python programming language and makes use of several programming libraries for processing of the data and illustration of the results as well for providing a comprehensive user interface.

3.1. Software packages and APIs

3.1.1. Python

The programming language used for implementing MetExtract is Python 2.6 (https://www.python.org, last accessed 20th July 2014). Python is a scripting language, which allows for fast implementation, as no compilation of the source code is required. In terms of performance, the execution of Python code is slower compared to code from a compiled programming language (e.g. C++).

3.1.2. PyQT

To provide the user with a comprehensive graphical user interface (GUI), PyQT 4.8 as well as the QT Designer, were used (http://www.riverbankcomputing.com, last accessed 20th July 2014).

3.1.3. MatPlotLib

MatPlotlib, which is a comprehensive plotting framework for the Python programming language [89], is used for illustration of the processed results. It can be integrated in PyQT created GUIs and supports a variety of different plot types (e.g. line and scatter plots).

3.1.4. Multiprocessing

Parallelisation of data processing in metabolomics can be achieved with little effort, since, in a first step, individual LC-HRMS measurements are processed independently of each other. For this functionality the Python package multiprocessing (method: imap_unordered) was

used (https://docs.python.org/2.6/library/multiprocessing.html, last accessed 20th July 2014). It allows defining a method *G* that is evaluated for *n* input parameters (e.g. LC-HRMS measurements). When the calculations are started, the package automatically distributes *m* calculations to *x* threads at a time. If more measurements than CPU cores need to be performed ($n>x$) the package processes a measurement once a previously started one has finished. Thus the CPU cores of a computer are being used most efficiently.

3.1.5. LC-HRMS data import

MetExtract uses the open-data format mzXML [90, 91] for importing centroid LC-HRMS data. Several conversion programs are freely available for this task (e.g. ReAdW [92] or msConvert from the ProteoWizard package [93]).

3.1.6. RPy2 and R

The python package RPy2 (http://rpy.sourceforge.net; last accessed 8th December 2014) was used for interfacing with R [94] during processing with MetExtract. The multiprocessing module was configured to automatically spawn an R-session for every thread that it executes.

3.1.7. MassSpecWavelet

The R-package MassSpecWavelet was used for chromatographic peak picking on EICs [95]. The package utilises wavelets for detecting peaks of variable width and intensity (used methods: getLocalMaximumCWT, identifyMajorPeaks, and tuneInPeakInfo; parameter 'scale' see Table 4 column 'Chromatographic peak width'). MetExtract uses the python package RPy2 to access the functionality of this package.

3.1.8. PTW – polynomial time warping

The R-package PTW (polynomial time warping; used method: ptw) is used for chromatographic alignment of several EICs in MetExtract [60]. It aligns *n* total ion chromatograms (TICs) or EICs by correcting slight shifts in the retention time with *n-1* polynoms ($p \in 1..x$). MetExtract uses the python package RPy2 to access the functionality of this package.

3.1.9. Statistical analysis with R

Statistical analysis including HCA, PCA, and univariate significance testing was performed with R [v. 3.1.0; 94].

3.1.10. Sum formula generation

To generate possible sum formulas for a given molecular mass *m* using the elements *E* (e.g. C, H, O, N, P, S), two methods were implemented:

- **Complete enumeration (CE)**
  This method generates all possible sum formulas consisting of the elements *E* for a mass *m*. CE does not verify the generated sum formulas for chemical logic or consistency (e.g. correct number of valence electrons) and also generates sum formulas, which are highly unlikely or even impossible (e.g. $C_{28}H_1$).

- **Seven golden rules (SGR, [24])**
  The SGR method for sum formula generation uses heuristic rules obtained from exhaustive metabolite database analysis and defines certain criteria (i.e. rules), which limit the search space for possible sum formulas for a given mass *m*. Thus, this method only generates plausible sum formulas and omits such which are unlikely or impossible. Furthermore, this method also checks for chemical logic or consistency. Consequently, sum formulas generated with SGR for a mass *m* are always a subset of the sum formulas generated with CE for the same mass *m*.

For sum formula generation only the principal isotopes of each element were allowed. The mass of the respective isotopes were obtained from the National Institute of Standards and Technology (NIST; http://physics.nist.gov/cgi-bin/Compositions/stand-alone.pl; last accessed 22nd December 2014)

3.2. Biological experiments and MetExtract settings used for data processing

The implemented software was used for automatic data processing in several untargeted metabolomics experiments that are summarised in Table 1 (publications #2-5; manuscript in preparation). Moreover, Table 2 lists the labelling setup used in the performed experiments, Table 3 summarises the most important analytical parameters of the utilised LC-HRMS platforms, and Table 4 specifies the data processing parameter settings used for evaluation with MetExtract.

3.2.1. Determination of labelling-isotope enrichment

The enrichment E with the respective labelling-isotope used in the labelled biological material (Table 2 column "Labelling enrichment") was determined as following (separately for each performed experiment):

1. Several (1-5) highly abundant feature pairs were manually searched for in a biological sample using TOPPView [96].
2. Using the determined number of carbon atoms a as well as the observed ratio o=M'+1/M', which implies one substation of ^{13}C with ^{12}C, the enrichment degree E was calculated using Equation 2.

$$E(a,s,o) = \frac{\binom{a}{s}^{\frac{1}{s}}}{\binom{a}{s}^{\frac{1}{s}} + o^{\frac{1}{s}}} \quad (2)$$

Table 1 | Overview of biological experiments that were processed with MetExtract.

Experiment / Publication	Experiment summary	Biological conditions	Biological replicates
DON in wheat / #2	This study was performed to investigate the detoxification mechanisms of wheat when stressed with the mycotoxin DON and to find previously unknown detoxification products of DON.	• Wheat ears treated with DON	5 (1 used for publication)
F. graminearum / #3	These biological experiments were preliminary aimed at demonstrating the benefits of SIL assisted untargeted metabolomics approaches using the developed workflows and MetExtract. Furthermore, they served as pre-trials for further biological experiments that are aimed at investigating the metabolome of the respective biological systems.	• Wildtype (PH1) • Mutant (*ΔTri5*) • Aggregate sample	6; 1 technical replicate of aggregate sample
Wheat / #3		• Susceptible (Remus) • Resistant (CM-82036)	5
Maize / #3		• Maize (CO354)	3
Phenylalanine in grape-berries / #4	This study was performed to gain insight into the metabolic changes in grape-berries exposed to heat (e.g. the warm climate of California). The focus was put on phenylalanine-containing metabolites (e.g. flavonoids), since many of these are associated with wine quality (colour and taste).	• Heat-stressed (45°C) • Control (25°C)	8
Phenylalanine in wheat cell suspension cultures / #5	The biological experiments were realized as a pre-trial to study if the wheat cell cultures metabolised the provided labelled phenylalanine. Furthermore, the biological workflow for tracer-fate studies was presented using the acquired LC-HRMS data.	• Treatment with DON • Control	3
T2 / HT2 in wheat / barley / In preparation	This study was performed to investigate the detoxification mechanisms of wheat and barley when stressed with the mycotoxins T2 and HT2 and to find previously unknown detoxification products.	• Wheat / Barley treated with T2 / HT2	3

Table 2 | Labelling setups used in the biological experiments presented in this work.

Experiment / Publication	Experiment Type	Tracer type	Labelling element	Labelling-isotope	Labelling enrichment	Parallel cultivation	Partly labelled tracer
DON in wheat / #2	Tracer-fate study	Exogenous	Carbon	^{13}C	99.5%	-	No
F. graminearum / #3	Full metabolome labelling	-	Carbon (Nitrogen)	^{13}C (^{15}N)	99.5% (99.5%)	Yes	-
Wheat / #3	Full metabolome labelling	-	Carbon	^{13}C	97.5%	No	-
Maize / #3	Full metabolome labelling	-	Carbon	^{13}C	97.5%	No	-
Phenylalanine in grape-berries / #4	Tracer-fate study	Endogenous	Carbon	^{13}C	98.8%	-	Yes
Phenylalanine in wheat cell suspen-sion cultures / #5	Tracer-fate study	Endogenous	Carbon	^{13}C	99.1%	-	No
T2 / HT2 in wheat / barley / In preparation	Tracer-fate study	Exogenous	Carbon	^{13}C	99.3%	-	No

Table 3 | Analytical instruments used for LC-HRMS analysis in the presented experiments.

Experiment / Publication	LC instrument type	LC manufacturer	HRMS instrument type	HRMS manufacturer	Ionisation polarity	Use of fast polarity switching
DON in wheat / #2	HPLC	Accela™	LTQ-Orbitrap XL	Thermo Fisher™	Positive	-
F. graminearum / #3	HPLC	Accela™	LTQ-Orbitrap XL	Thermo Fisher™	Positive	-
Wheat / #3	HPLC	Accela™	LTQ-Orbitrap XL	Thermo Fisher™	Positive	-
Maize / #3	HPLC	Accela™	LTQ-Orbitrap XL	Thermo Fisher™	Positive	-
Phenylalanine in grape-berries / #4	HPLC-Chip	Agilent Technologies™	Quadupole Time-of-Flight	Agilent Technologies™	Positive (Negative)	No
Phenylalanine in wheat cell suspension cultures /#5	HPLC	Dionex™	Orbitrap Exactive plus	Thermo Fisher™	Positive / Negative	Yes
T2 / HT2 in wheat / barley / In preparation	UHPLC	Agilent Technologies™	Quadrupole Time-of-Flight	Agilent Technologies™	Positive	No

Table 4 | Parameter settings used for automated data processing in the presented experiments.

Experiment / Publication	Mass precision	Minimum abundance	Verified isotopologs	Maximum isotopolog ratio error	Accepted M:M' ratio	*m/z* clustering	EIC width	Chromato-graphic peak width	Minimum correlation	Feature pair bracketing
Processing step(s) in MetExtract	1.1, 4, 5	1.2	1.3, 1.4	1.3, 1.4	1.5	2	3, 9	3, 9	3	8
DON in wheat / #2	± 4 ppm	5,000 counts	M+1, M'-1	± 20%	0.91 ± 0.4	8 ppm	± 5 ppm	10-40 sec	0.5	-
F. graminearum / #3	± 4 ppm	5,000 counts	M+1, M'-1	± 20%	-	8 ppm	± 5 ppm	10-40 sec	0.5	± 8 ppm
Wheat / #3	± 4 ppm	5,000 counts	M+1, M'-1	± 20%	-	8 ppm	± 5 ppm	10-40 sec	0.5	± 8 ppm
Maize / #3	± 4 ppm	5,000 counts	M+1, M'-1	± 20%	-	8 ppm	± 5 ppm	10-40 sec	0.5	± 8 ppm
Phenylalanine in grape-berries / #4	± 20 ppm	100 counts	M+1, M'-1	± 20%	-	40 ppm	± 40 ppm	15-40 sec	0.5	± 40 ppm
Phenylalanine in wheat cell suspen-sion cultures / #5	± 3 ppm	50,000 counts	M+1, M'-1	± 20%	-	8 ppm	± 5 ppm	10-40 sec	0.75	± 8 ppm
T2 / HT2 in wheat / barley / In preparation	± 5 ppm	20,000 counts	M+1, M'-1	± 20%	1.0 ± 0.5	10 ppm	± 10 ppm	10-40 sec	0.75	± 10 ppm

4. Results

MetExtract, a software tool for the processing of LC-HRMS derived metabolomics data in SIL assisted untargeted metabolomics and tracer-fate applications, was developed and implemented. The software detects and verifies feature pairs using the highly unique isotope patterns of native and labelled metabolite-derived ions, which enables the holistic detection of only truly biological relevant metabolites in full metabolome labelling approaches or biotransformation products that are descendants of the tracer substance under investigation. Additionally, the determination of the number of atoms of the labelling-isotope allows improved metabolite annotation compared to labelling-free approaches, which rely on the relative abundances of different isotopologs. Using appropriate experimental and sample preparation protocols, metabolome- and experiment-wide internal standardisation is achieved, which improves comparative metabolite quantification. Furthermore, in tracer-fate studies MetExtract specifically detects all known and unknown biotransformation products derived from studied tracer substances.

In the following the developed and implemented data processing steps are summarised. These steps are depicted in Figure 8.

4.1. Mass spectrum based data processing

MetExtract first inspects each LC-HRMS data-file separately for feature pairs derived from native and labelled metabolites. To this end, putative MS pairs are detected with a two-dimensional data-filtering algorithm, which initially operates on a mass spectrum level and then uses the detected MS pairs for the evaluation of the data in the chromatographic domain (Figure 9).

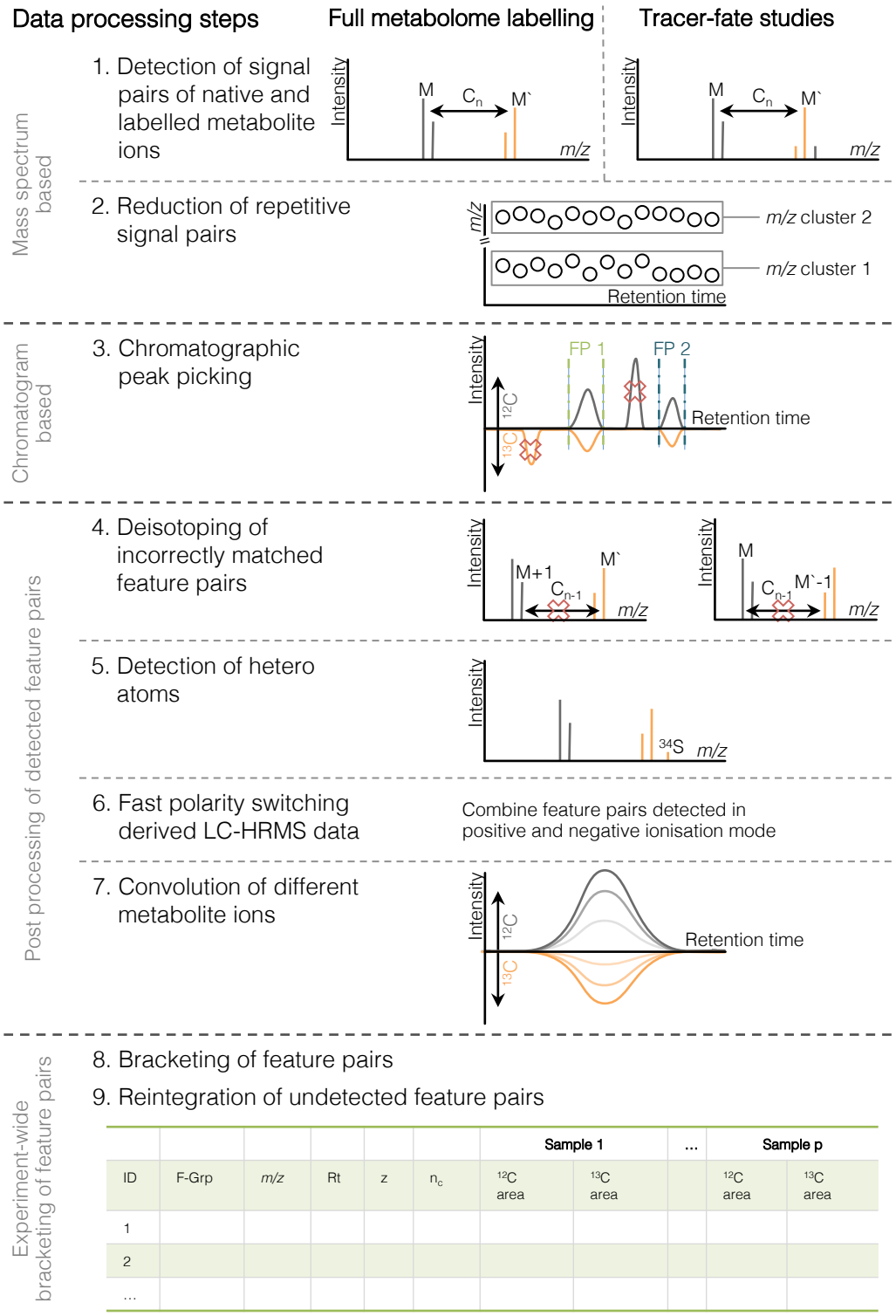


Figure 8 | Illustration of the data processing steps implemented in MetExtract: Initially signal pairs are detected on an MS spectrum level ("Mass spectrum based data processing steps"; steps 1 & 2). Subsequently, these signal pairs are used in the next step to separate different homologs ("Chromatogram based data processing steps"; step 3). Extracted feature pairs are further annotated and convoluted ("Post processing of detected feature pairs data processing steps"; steps 4-7). Finally, all feature pairs found in the individual LC-HRMS measurements are bracketed ("Experiment-wide bracketing of feature pairs data processing steps"; steps 8 & 9). Figure adapted from publications #3 and #5.

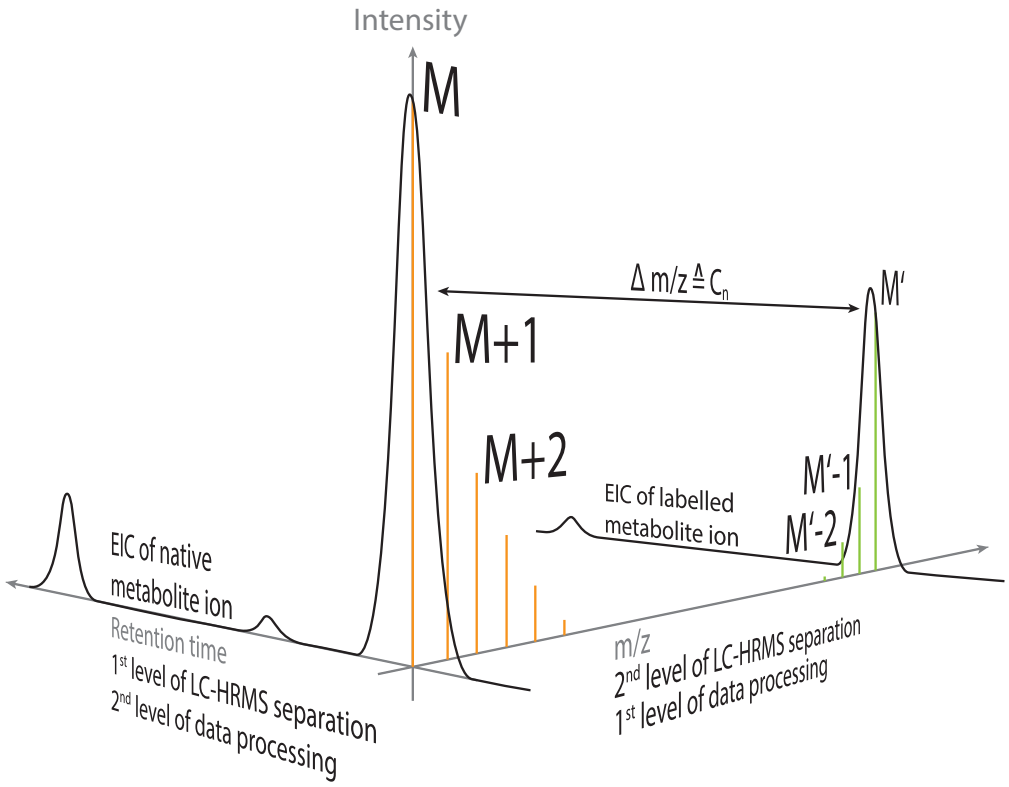


Figure 9 | Schematic representation of LC-HRMS data and two-dimensional feature pair detection with MetExtract. First, the characteristic isotope patterns of native (|M, M+1, ...|; orange mass peaks) and ^{13}C-labelled (|M', M'-1, ...|; green mass peaks) metabolite-derived ions are searched for in each recorded MS spectrum. Then, the detected MS signal pairs are used to separate different isomers in the chromatographic domain of the acquired LC-HRMS data. Chromatographic peaks present either for native or labelled metabolite ions but not the respective other form are discarded.

4.1.1. Data processing step 1: Detection of signal pairs of native and labelled metabolite ions

To detect isotope patterns derived from native and ^{13}C-labelled metabolites, MetExtract uses a brute force approach. It iterates over all MS scans present in a LC-HRMS measurement and assumes each mass peak in a scan to be a valid, monoisotopic mass peak M of a putative metabolite ion. This assumption is then verified with the following criteria:

1. **Labelled metabolite analogous**
   Each valid mass peak M must have a corresponding mass peak M'. Their m/z difference must correspond to a multiple of the mass difference between ^{12}C and ^{13}C, which is 1.00335. Since the number of carbon atoms in a metabolite-derived ion is not known afore, a certain range of integer values corresponding to a variable number of carbon atoms is tested. For each integer value i, which is assumed to be the number of labelling-isotopes in M', and charge z, M' is calculated as an

offset to M (i.e. M'=M+i*1.00335/z). A mass peak is considered to represent a valid M', if a mass peak with the calculated m/z value is present within a certain error windows (in ppm) in the same mass spectrum as M. Subsequently, the mass peak M is annotated with the respective number of carbon atoms i and the charge state z, and the next verification step is performed.

2. **Minimum abundance**
   M and M' peak pairs that are below a predefined, minimum relative intensity are discarded. If the intensities for both mass peaks are above the threshold, the verification continues.

3. **SIL-derived isotope pattern verification**
   The number of determined carbon atoms per metabolite ion is verified using the observed isotope pattern of the labelled metabolite-derived ion. To this end, the theoretical ratio t=P(i, 1, p), with i denoting for the determined number of labelling-isotopes and p for the isotopic enrichment with ^{13}C, is compared to the observed ratio o=M'-1/M'. For this, first the mass peak M'-1 with the m/z value M'-1=M+(i-1)*1.00335/z needs to be present in the same scan. If additionally the theoretical ratio t and observed ratio o are approximately the same ($|t-o| \leq$ threshold), this isotope abundance check is passed.

4. **Native isotope pattern verification**
   This data processing step verifies the carbon-isotope pattern of the native metabolite ion. For this, first the mass peak M+1 with the m/z value M+1=M+1.00335/z is searched for. If found, the carbon-isotopolog M+1 is tested against the determined number of labelling-isotopes. Depending on the experiment type (either a full metabolome labelling experiment or a tracer-fate study), different criteria are asserted:

   a. **Full metabolome labelling experiment**
      In a full metabolome labelling experiment all carbon atoms are replaced with ^{13}C during the SIL process resulting in globally labelled metabolites. The ratio M+1/M is derived from all carbon atoms present in the metabolite-derived ion and can be tested likewise the M'-1/M' ratio verification in the previous verification step. Thus, the theoretical ratio t=P(i, 1, 0.9893) of the determined number of carbon atoms i is compared to the observed ratio o=M+1/M. If the theoretical ratio t and observed ratio o are approximately the same ($|t-o| \leq$ threshold), this isotope abundance check is passed.

b. **Tracer-fate studies**

Biotransformation products, which are a conjugation of a ^{13}C-labelled tracer and a native moiety, are generally not globally labelled. Any moiety that originates from the studied biological system or a non-labelled part of the studied tracer substance is present as a native form resulting in partly labelled metabolites. Thus, the ratio M+1/M is not only derived from the determined number of ^{13}C-isotopes but the total number of carbon atoms present in this biotransformation product. Consequently, the corrected, observed ratio o=(M+1/M)-(M'+1/M') is calculated. If M'+1 with the m/z value M+(i+1)*1.00335/z is not found, it is considered that the biotransformation product does not contain any native moiety and thus the observed ratio is corrected to o=M+1/M. The observed ratio o is then compared to t=P(i, 1, 0.9893). If these two ratios are approximately the same ($|t-o| \leq$ threshold), this isotope abundance check is passed.

5. **Exogenous tracers (optional check for exogenous tracer substances)**

Exogenous tracers need to be supplied as native and labelled metabolites during the biological experiment (Figure 4). Consequently, the ratio a=M:M' of the studied tracer is known. As native and labelled molecules of the same substance are metabolised to an equal extend in biological systems, tracer-derived biotransformation products also have to have the same ratio as the applied, pure tracer. Thus, only if the ratio o=M:M' for the observed abundances of a putative biotransformation product approximately match with the applied ratio a ($|a-o| \leq$ threshold) this optional verification step for an exogenous tracers is passed. For full metabolome labelling and endogenous tracer-fate studies this check is skipped.

The result of this signal pair detection and elaborate verification is a list of mass peak pairs that putatively originate from native and labelled ions of the same metabolites. Each entry in the list contains an observed, monoisotopic m/z value of M, a scan number in which this signal pair was detected, the number of labelling-isotopes, a charge state z, and the ratio M:M'. Consequently, this lists also includes matched isotopologs that are incorrectly assumed to be either monoisotopic or uniformly/consistently labelled metabolite ions (e.g. M+1 paired with M'; M paired with M'-1). These incorrect pairings are removed later in data processing step 4.

4.1.2. Data processing step 2: Reduction of repetitive signal pairs

Typically, in LC-HRMS analysis of a biological sample metabolite molecules elute for several seconds from the chromatographic column and therefore are recorded in consecutive MS scans. Different isomers, which are metabolites with the same chemical formula but a different

structure, also have identical m/z values. To reduce the number of these redundant mass peak pairs, they are grouped using hierarchical clustering with Euclidean distance. First, the hierarchical tree is calculated from all detected signal pairs. In the second step, the generated dendrogram is processed starting from the single top cluster, which contains all signal pairs. The cluster and following sub-clusters are branched if the m/z value difference between the highest and lowest signal pair is higher than a pre-set threshold in ppm, which is typically a multiple of the used instrument's mass accuracy. If this mass difference in a sub-cluster is below the threshold, the branching is stopped. A cluster, which is not further branched, represents an aggregation of similar signal pairs (i.e. mean of several MS scans of a chromatographic peak and/or chromatographically separated isomers).
Hierarchical clustering and branching is performed separately for each labelling-isotope number and charge states. Consequently, it reduces the number of similar signal pairs detected in the previous step and also reports a mean m/z value for each remaining sub-cluster.

4.2. Chromatogram based data processing

After targets in form of clustered signal pairs have been detected on a mass spectrum level, they are further processed with the aim to separate different isomers with the same chemical formula but a different structure and thus different chromatographic peaks (Figure 8). To this end, the information from the chromatographic separation – the retention time – is used.

4.2.1. Data processing step 3: Chromatographic peak picking

Native and ^{13}C-labelled metabolite-derived ions are chromatographically not separated. Thus, all ions derived from native and labelled metabolites must show the same chromatographic peak profiles in the respective EICs of M and M'. This allows separating different isomers, but also verifies if both the native and labelled ions of a metabolite are present in the same LC-HRMS data.
To identify SIL-derived feature pairs, chromatographic peaks are searched for in the EICs of M and M' for each detected signal pair cluster using the MassSpecWavelet package of Du [95]. Only coeluting chromatographic peaks (i.e. have very close peak centers; ± number of scans) present in both EICs of a putative feature pair are kept. These chromatographic peak pairs are further verified with a check for highly similar peak profiles using the Pearson correlation coefficient. Only if the correlation of the two chromatographic peaks between their respective peak-borders as determined by the peak picking exceeds a pre-set threshold, they are considered to perfectly coelute and thus definitely represent a SIL-derived feature pair. Chromatographic peaks with a low correlation or present only in one of the two EICs are discarded from the results.

After these data processing steps, each such detected feature pair corresponds to a unique metabolite ion derived from a native and ^{13}C-labelled metabolite. Such feature pairs have a mean m/z value of M, a retention time *Rt*, the determined number of ^{13}C-labelling-isotopes, a charge z, and two peak areas for the chromatographic peaks of M and M'.

4.3. Post processing of detected feature pairs

Each feature pair detected in the previous, two-dimensional data filtering steps represents a single ion derived from native and labelled metabolite ions (Figure 8). In the following steps, they are further verified and annotated.

4.3.1. Data processing step 4: Deisotoping of incorrectly matched feature pairs

The previous analysis steps do not consider that different isotopologs of both the native and labelled metabolite-derived ions may be matched incorrectly (e.g. M+1 with M'; M+1 with M'-1). To remove these invalid feature pairs, MetExtract searches for further feature pairs, which 1) elute at approximately the same retention time, 2) have a positive m/z offset corresponding to one or several labelling-isotopes, and/or 3) a reduced number of labelling-isotopes compared over the currently verified feature pair. Feature pairs matching these criteria are considered incorrect pairings and removed from the results. This simple, but essential procedure efficiently removes incorrect pairings and only keeps those feature pairs, which represent correct pairings of M and M' ions.

4.3.2. Data processing step 5: Detection of heteroatoms

Some hetero-elements (e.g. sulphur, chloride, iron) have naturally occurring stable isotopes that can be observed in LC-HRMS data since their relative abundances compared to the principal isotope of the respective element are quite high (e.g. ^{34}S with 4.4% abundance relative to ^{32}S). Thus, metabolite ions that contain atoms of such elements show distinct isotope patterns in the LC-HRMS data given a high abundance of the respective monoisotopic or uniformly/consistently labelled metabolite ions and a sufficient MS resolution. To this end, the theoretical m/z value of a heteroatom-isotopolog is calculated and searched for. In case of a mass increase of the less abundant stable isotope compared to the element's principal isotope (e.g. sulphur-34 with +1.9958 u relative to ^{32}S), the search is performed starting from M' (e.g. for sulphur-34 at M'+1.9958/z), while for isotopes with a mass decrease compared to the element's principal isotope (e.g. ^{54}Fe with -1.9953 u relative to ^{56}Fe) the search is performed starting from M (e.g. for ^{54}Fe at M-1.9953/z). This calculated m/z is searched for in every mass spectrum that is within the chromatographic peak of the respective feature pair. If the ob-

served mean ratio of the heteroatom isotopolog approximately corresponds to the theoretical ratio observed for *n* instances of such a heteroatom, the feature pair is annotated with this putative heteroatom.

4.3.3. Data processing step 6: Fast polarity switching derived LC-HRMS data

LC-HRMS data utilising fast polarity switching contains mass spectra derived from the positive and negative ionisation mode. As a result, the same metabolite may be recorded as different ions specific for the positive and negative ionisation mode (e.g. $[M+H]^+$, $[M-H]^-$). In this respect all of the previously described steps (1-5) for feature pair detection are separately executed for the positive and negative ionisation modes resulting in two distinct feature pair lists for the positive and negative ionisation mode. In the following steps, these lists are combined.

4.3.4. Data processing step 7: Convolution of different metabolite ions

During ionisation of the uncharged metabolites a variety of different ion species may be formed. Thus, the next step in MetExtract is the convolution of different feature pairs into feature groups each representing a unique metabolite in the analysed sample. For this, the chromatographic peak shapes of closely coeluting feature pairs are compared. Two such pairs that show a high Pearson correlation coefficient are linked. All linked metabolic features represent a single feature group and thus are very likely to originate from the same metabolite. Furthermore, if fast polarity switching is utilised, polarity-specific ion species also show perfect coelution and no chromatographic shift between the two polarity modes can be observed. Thus, even feature pairs from the positive and negative ionisation mode can be successfully convoluted into feature groups (e.g. Figures 6 and 7 in Chapter 2).

After convolution, feature groups are inspected for relationships between the different feature pairs (e.g. different ion species or in-source fragments). For this, all pairs in a feature group are compared pairwise. Initially it is checked, whether the relationship between two feature pairs is explained by two common ion species. If such a combination (e.g. $[M+H]^+$ and $[M+Na]^+$) is found, and the detected number of labelling isotopes is identical in both feature pairs, these two are annotated with their respective ion species and the mass for the native, monoisotopic molecule is calculated. If ion species annotation fails, sum formulas for putative neutral losses are calculated using the m/z value difference and the determined number of ^{13}C-isotopes of the two feature pairs [24]. The feature pair with the higher m/z value is then annotated with the generated sum formula, while the feature pair with the lower mass is annotated with the loss of the respective atoms. Only in the special case that two feature pairs derived from positive and negative ionisation modes have a m/z difference of

$2*1.007276/z=2.014552/z$, which corresponds to a protonated ion ($[M+H]^+$) in the positive ionisation mode and a deprotonated ion ($[M+H]^-$) in the negative ionisation mode, their relationship is directly annotated with these two ion species.

4.4. Experiment wide bracketing of feature pairs across multiple data-files

Metabolomics experiments usually consist of many LC-HRMS measurements (e.g. biological and technical replicates, different experimental conditions) (Figure 8). To compare their metabolic states, the same metabolites need to be tracked across all measurements. Slight differences across the LC-HRMS measurements include mass accuracy or retention time shift.

4.4.1. Data processing step 8: Bracketing of feature pairs

To account for minor chromatographic retention time shifts, usually a chromatographic alignment is performed. In this work the R-package PTW [60] was utilised for this task. In MetExtract, EICs of the ions M' are used for the alignment step, because, according to the established full metabolome labelling workflow, they originate from pooled, labelled sample material, which is identical with respect to its metabolic constitution and relative abundance in each sample.

In MetExtract, the term bracketing describes the process of summarising the same feature pairs across different measurements. This is accomplished using the m/z value of M, the charge state z, the mean retention time Rt, the ionisation mode, and the number of labelling-isotopes of feature pairs detected in the individual LC-HRMS measurements. To this end, first all feature pairs with the same ionisation mode, the same number of labelling-isotopes, and the same charge state are clustered using HCA on their m/z value. Again, the generated dendrogram is split until sub-clusters remain, where all masses are within a multiple of the used instrument's mass accuracy. Each sub-cluster is then clustered using the retention times of all feature pairs in the sub-cluster. This retention time dendrogram is then branched similar to the m/z value tree. Sub-clusters, which have less retention time deviation than a typical chromatographic peak, are kept and denote for a bracketed feature pair.

4.4.2. Data processing step 9: Reintegration of undetected feature pairs

After bracketing, those feature pairs, which are not consistently found in all LC-HRMS measurements, are searched for in those samples, in which MetExtract did not detect them (e.g. too low abundance of necessary isotopologs in biological conditions). To this end, the

EICs of bracketed feature pairs are extracted in those samples, in which they were not detected. Chromatographic peak peaking is then performed with the same settings as in step 3 in those EICs and detected chromatographic peaks of M and M' are recorded. The isotopolog information, which would be necessary to detect a feature pair, is, however, not verified. Consequently, this step is a semi-targeted search using feature pairs found in other samples and cannot detect new feature pairs. This reintegration is especially helpful during subsequent statistical analysis (e.g. detecting differentially abundant metabolites).

4.5. Data processing output

MetExtract stores the results of automated data processing in various forms. Both, the results of individual LC-HRMS data-files as well as a data matrix containing detected and bracketed feature pairs from all samples are stored. In the following these results are summarised:

For each processed LC-HRMS data-file, which represents a single sample in the metabolomics experiment, MetExtract creates the following results:

- **Feature pair list**
  This tab-separated values file (.tsv) contains all feature pairs detected in a certain sample. Each feature pair is annotated with the m/z value of M, its retention time Rt, the determined number of atoms of the isotope used for labelling, the relative abundances of M and M', the ionisation mode in which this feature pair was detected, its charge z, and the metabolite group to which it was assigned.

- **Feature pair PDF**
  Apart from the list with the detected and convoluted feature pairs, MetExtract also renders a PDF file that contains graphical illustrations of the results. The first pages document the data processing parameter settings for generating these results. Then, on each page one feature pair is depicted with the same information as in the tsv-list. Additionally, the EICs of M and M' and a mass spectrum showing the different isotopologs of the feature pair are drawn. Furthermore, for each feature group all EICs of its feature pairs are overlaid and the feature group annotation is listed.

- **mzXML file**
  Besides the processed results lists, a new mzXML file, which contains only the different isotopologs (i.e. |M, M+1, … M'-1, M', M'+1…|) of the detected feature pairs, is created by MetExtract. However, this representation does not include any meta-information (e.g. determined number of atoms of the used labelling-isotope, relative metabolite abundances) of the detected feature pairs.

Besides the result files created for individual LC-HRMS data-files, MetExtract also stores the results of all samples in an experiment in form of a feature pair data matrix. All bracketed

results detected across the different LC-HRMS measurements are stored as a tab separated values file (.tsv). Each line represents a single feature pair and is annotated with a mean m/z value of M, a mean retention time Rt, the determined number of atoms of the used labelling-isotope, the respective ionisation mode in which it was detected, and its charge z. Furthermore, for each LC-HRMS measurement the abundance of M and M' as well as the determined feature group are stored separately.

Moreover, MetExtract offers a comprehensive user interface to process and review the results of individual LC-HRMS data-files. Illustrations of the processed results include a two-dimensional feature map (retention time vs. m/z of detected feature pairs), normalised EICs of individual feature pairs and convoluted feature groups, annotated mass spectra, as well as the used data processing parameter settings.

4.6. Applications of MetExtract

The workflows for full metabolome labelling and tracer-fate studies in untargeted metabolomics experiments and the implemented software tool were applied in several analytical and biological studies. Results of these studies were published in peer-reviewed journals and are reprinted in Part II (publications #2-#5). In the following the main findings of each study are briefly summarised.

The software tool MetExtract was used for studying the metabolic fate of the exogenous mycotoxin deoxynivalenol (DON) in wheat plants (publication #2; SFB project "Fusarium"). This mycotoxin is a virulence factor of the fungi *F. graminearum* and is transformed to less toxic metabolites by wheat to counteract fungal infection [97]. A total of 9 different DON detoxification metabolites were found, 5 of which have not been known previously.
In a similar experiment, the metabolic fate of the mycotoxins T2 and HT2 is currently investigated in wheat and barley (collaboration with Marc Lemmens; manuscript in preparation). There, several yet unknown and uncharacterised biotransformation products have been detected.

Another application of MetExtract details the biological and analytical workflows for full metabolome labelling approaches (publication #3; SFB project "Fusarium"). *F. graminearum* was cultivated on native and globally ^{13}C-labelled glucose. Furthermore, native and commercially acquired ^{13}C-labelled wheat and maize plants were analysed with LC-HRMS. These datasets were then facilitated for demonstrating the advantages of SIL in untargeted metabolomics approaches with the aim of detecting as many metabolites as possible. More specifically it was shown, that only true biological metabolites are detected with the aid of MetExtract and that experiment- and metabolome-wide internal standardisation greatly enhanced the analytical precision and aided in statistical analysis of the gained metabolomics data. Additionally, these biological applications provided first insights into the respective metabolomes and served as pilot studies for further investigation of fungi and plant interactions.

MetExtract was applied to study the metabolic fate of the endogenous amino acid phenylalanine in grape-berries (publication #4; cooperation with Alexander Chassy and Andrew Waterhouse from the Waterhouse Lab; http://waterhouse.ucdavis.edu; last accessed 23rd October 2014). The Waterhouse group investigates why wine, which is produced in California, has diminished quality (less taste and colour) compared to wine produced from vineyards in other geographical areas. It is believed that the warm climate of California reduces the wine's flavonoid content, which greatly contribute to the reduction of its taste and colour [98]. To study this effect, the group chose to investigate flavonoid metabolites using a tracer-fate approach. To this end, the metabolic fate of phenylalanine and its derived secondary metabolites were studied in grape-berries during their ripening. 16 samples, 8 heat-stressed and 8 controls, were analysed and 63 metabolites were detected, many of which could be identified as flavonoids. Several metabolites showed increased abundance in the heat stressed samples compared to the control grape-berries. Hence these represent targets for further investigation of heat-stress in grape-berries.

In the most recent tracer-fate application study, MetExtract was used to investigate the metabolic fate of phenylalanine in wheat cell suspension cultures, because many phenylalanine-derived metabolites in plants are known to have anti-fungal activity [99] (publication #5; SFB project "Fusarium"). In this experiment, the LC-HRMS analysis was performed with an Orbitrap Elite plus instrument, which is capable of fast polarity switching. A total of 139 different metabolites were found. 49 of those metabolites showed feature pairs in both ionisation modes. For 19 of the detected metabolites, the uncharged, monoisotopic mass could only be calculated based on complementary feature pairs detected in both ionisation modes, which demonstrates the great benefit of fast polarity switching derived LC-HRMS data.

5. Discussion

In the following, the main challenges addressed by the developed workflows in combination with the custom-designed MetExtract software are presented and discussed with selected biological experiments.

5.1. Requirements and limitations of MetExtract

MetExtract was designed to support different LC-HRMS platforms (e.g. Orbitrap, Time-of-Flight) and isotopes used for labelling (e.g. ^{13}C, ^{15}N, ^{34}S). The developed workflow and the implemented software, however, suggest ^{13}C as the main labelling-isotope since carbon is present in virtually any metabolite, which enables the simultaneous detection of as many metabolites as possible in untargeted metabolomics applications. On the contrary, if the labelling is performed with ^{15}N or ^{34}S, only such metabolites containing at least one atom of the respective element, may be detected.

With respect to isotopic enrichment, the software requires distinct isotope patterns derived from native and labelled metabolite ions. It supports full metabolome labelling experiments and tracer-fate studies with labelled metabolites generally having an isotopic enrichment with the used labelling-isotope of at least 97%. It is, however, not possible to state these limitations absolutely as the labelling-parameters vary in any experiment and need to be evaluated for each experiment separately. Moreover, the isotope patterns of native and labelled metabolites must not overlap. In labelled metabolites that meet these criteria the principal isotopologs M and M' of the native and labelled metabolites can be determined, which allows calculating the exact number of labelling-isotopes in each detected feature pair and metabolite. In case of ^{13}C-labelling, only metabolites having at least 5 carbon atoms can be detected. If the labelling is performed with an isotope other than ^{13}C, the LC-HRMS instrument must be operated in a way that allows separating the isotope patterns of carbon and the isotope used for labelling, as the current version of MetExtract does not support deconvolution of overlapping isotope patterns from different elements.
For tracer-fate studies the tracer-derived part of any biotransformation products must also fulfil these criteria. Thus, MetExtract is mainly suited to study the metabolic fate of such compounds, which are not intensively fragmented (e.g. drugs, toxins, or secondary endogenous metabolites). With respect to chromatographic separation, the software can detect chromatographic peaks of variable width (typically 10-40 seconds with a minimum of one MS scan per second). For metabolome and experiment-wide internal standardisation, MetExtract requires that the absolute amount of labelled material must be identical in all studied samples. Furthermore, it must be possible to export the acquired raw-LC-HRMS data into the mzXML format.

5.2. Contrasts of MetExtract to X^{13}CMS

Recently Huang and colleagues released the X^{13}CMS workflow [80]. This workflow was designed to detected changes in the isotopic patterns of metabolites, which subsequently allows annotating those metabolites affected by the provided isotopically labelled substance (e.g. ^{13}C-labelled glucose). Complementary, X^{13}CMS also reports metabolites, which were not affected by the investigated substance.

The biological experiments required for the X^{13}CMS and MetExtract tools are fundamentally different. X^{13}CMS requires native and isotopically labelled samples to be analysed independently for each experimental condition. On the contrary, MetExtract requires LC-HRMS data from mixtures of native and isotopically labelled samples for each experimental condition.

During feature detection, X^{13}CMS utilises the XCMS algorithm to detect features (i.e. chromatographic peaks) present in the LC-HRMS data. Subsequently, the tool convolutes different isotopolog features originating from the same ion species of a compound to search for altered isotope distributions that are a result of the treatment with the isotopically labelled tracer substance. X^{13}CMS then reports all detected features either affected by the studied tracer or not. In contrary, MetExtract utilises the isotopic information first to detect putative targets that are subsequently used for chromatographic peak picking. Thus, MetExtract only extracts such feature pairs, which originate from native and labelled metabolites and discards all metabolites, which do not show a labelled analogue.

After feature detection X^{13}CMS uses differential abundance analysis to report those features that show statistically significant differences between the studied experimental conditions. Consequently, X^{13}CMS does not support experiment- and metabolome-wide internal standardisation, since it requires non-pooled labelled sample material. On the other hand, MetExtract and the developed workflow for full metabolome labelling support experiment- and metabolome-wide internal standardisation and thus allow accounting for different matrix effects and LC-HRMS drifts across measurement batches.

X^{13}CMS does not require globally labelled sample material. As a consequence it does not support determining the total number of labelling-isotopes present in a metabolite-derived ion. On the contrary, MetExtract requires native and globally labelled sample material in full metabolome labelling experiments and thus reports the total number of labelling-isotopes in each detected metabolite-derived ion species.

Consequently, X^{13}CMS and MetExtract are designed for different purposes:

i. X^{13}CMS is used to elucidate the metabolic fate of certain tracer substances that are incorporate into many metabolites. Moreover, with X^{13}CMS differences in the flux of all tracer-derived metabolites can be detected between two or more experimental conditions. Moreover, it also allows extracting those metabolites from the

LC-HRMS data that are not affected by the isotopically labelled tracer substance under investigation.

ii. MetExtract is designed for the detection of all biological relevant metabolites in full metabolome labelling approaches and utilises pooled labelled sample material to improve relative quantification and subsequently performed statistical analysis. In tracer-fate approaches, MetExtract can also detect secondary metabolites that are derived from a certain tracer substance under investigation. Additionally, it allows determining the total number of labelling-isotopes in each found metabolite or biotransformation product, which greatly enhances their annotation and characterisation. Furthermore, using native and labelled metabolites, MetExtract efficiently discards any non-biological relevant artefacts and background signals.

In summary, X^{13}CMS and MetExtract are similar, as both require native and isotopically labelled sample material, but are different in respect to the biological questions asked. This highlights that both tools represent valuable contributions for untargeted metabolomics research and depending on the biological question(s) asked the experimental setup and data processing tools have to be chosen.
As a consequence of the different designs of X^{13}CMS and MetExtract, no comparative analysis on a dataset could be performed.

5.3. Feature pair detection, convolution and annotation

MetExtract detects truly biology-derived metabolic features present as native and labelled metabolite ions in acquired LC-HRMS data. Each such feature pair denoting an ion species of a certain metabolite is verified in three steps: 1) the distinct m/z difference between the monoisotopic and uniformly/consistently labelled isotopologs, which corresponds to the number of atoms of the used labelling-isotope, 2) the isotope pattern for both the native and labelled metabolite ions, and 3) a check for perfect coelution. These criteria provide a high degree of confidence that the detected metabolic features are truly derived from the biological system. Consequently, all contaminants and non-biology related substances and their ions are filtered out, as they in general do not satisfy those criteria. While verification step 1, which operates on the MS spectrum level, is prone to also detect false positives (e.g. pairings of noise or Fourier transform artefacts [100]), verification step 2 reduces the number of false positives significantly. Verification step 3 further removes incorrect pairings and confirms already detected m/z clusters. A subsequent deisotoping step then removes those feature pairs that originate from incorrectly paired isotopologs. Figures 6 and 7 in Chapter 2 illustrate 5 different feature pairs of a globally ^{13}C-labelled metabolite and a partly ^{13}C-labelled biotransformation product, which fulfil all those criteria and thus were automatically detected by MetExtract.

The developed workflows and MetExtract for detection of metabolites using SIL-derived isotope patterns are highly specific. Other substances and noise not meeting these criteria are

efficiently filtered, which is demonstrated in the publications #3 and #5. In both publications, background signals and non-labelled substances were simulated using native biological matrixes. Consequently, these samples contained no labelled metabolites at all. Analysis of the data with MetExtract resulted in a small number (on average less than 5 per analytical sample) of incorrectly detected feature pairs. Manual investigation of these false positives showed that most of them were incorrect pairings of Fourier transform artefacts [100, 101] of different metabolites, which happened by chance. Fourier transform artefacts are signals originating from a compound that is highly abundant. These signals have a characteristic m/z offset and usually a very low abundance, which is only a few percent of the actual metabolite-derived mass peak. For example, in the metabolomics data derived from the used LTQ-Orbitrap XL instrument (publication #3) an m/z shift of approximately 20 ppm with an abundance of less than 3% compared to the correctly recorded signal was observed. Such artefacts may also be present for native and labelled metabolite ions and their respective isotopolog signals and are thus detected as individual feature pairs by MetExtract. However, during metabolite convolution they are summarised into the same feature group since they show perfect coelution with the correct metabolite-derived signals.
A low number (1-5) of such Fourier transform artefacts were detected in the *F. graminearum* full metabolome labelling experiment (publication #3). Furthermore, also in publication #5 Fourier transform artefacts were the majority of incorrectly extracted feature pairs. For most of the incorrectly detected feature pairs the number of carbon atoms as determined by the isotopolog pairing was quite small (less than 15) while at the same time the m/z value of M was higher than expected for a metabolite having this few carbon atoms (higher than m/z 400), which is another indication that these feature pairs represent false positives.

In the experiments presented in this work, the degree of ^{13}C enrichment was always higher than 97%. Furthermore, only a few metabolites with more than 50 carbon atoms were detected. As a consequence, the signals M and M' were the most abundant ones among their respective carbon isotope patterns. Consequently, MetExtract correctly filtered all incorrect pairings of M+x and/or M'-x isotopologs and listed only the feature pairs with M and M'.

In summary, the unique isotope pattern of native and labelled metabolite-derived ions in LC-HRMS data is perfectly suited for the very efficient filtering of noise and background signals. In this respect, MetExtract automatically detects all truly biology-derived metabolites and biotransformation products with only a few false positives.

5.3.1. Improved feature pair annotation

In full metabolome labelling experiments all metabolites are globally labelled with a stable isotope (e.g. ^{13}C). This allows determining the total number of atoms of the isotope used for labelling in each detected feature pair. This information is especially helpful to annotate the unknown metabolites or to generate possible sum formulas [102].

In the following, the improved annotation of known and unknown metabolites with the exact number of carbon atoms (C_n) during sum formula generation is exemplified. Box 1 lists the steps used for sum formula generation and Table 5 shows the results for 4 randomly selected sum formulas as well as for the full metabolome labelling experiment of *F. graminearum* presented in publication #3.

Box 1 | Sum formula generation in full metabolome labelling experiments.

Sum formula generation: Possible sum formulas for a mass *m* were generated with i) CE and ii) SGR. For both methods a ± 5 ppm error between *m* and the mass of a generated sum formula was tolerated in the randomly selected sum formulas and a ± 3 ppm error was allowed for the *F. graminearum* derived feature pairs. Both methods were restricted to $C_{n\pm x}H_{0\text{-}130}O_{0\text{-}40}N_{0\text{-}10}P_{0\text{-}10}S_{0\text{-}10}$, however, SGR may have chosen to violate some of these elemental constraints given the heuristic nature of the algorithm.
Isotopolog uncertainty: To show the benefit of using the exact number of carbon atoms (C_n) during metabolite annotation, a 0 (exact number of carbon atoms), 3, or 10 carbon atom inaccuracy was used ($C_{n\pm x}$).
Simulation: Four randomly selected sum formulas were used. For each, putative sum formulas were generated with CE and SGR independently. This was done for the exact number of carbon atoms (C_n) as well as the simulated uncertainties. Furthermore, sum formulas were also generated for all feature pairs detected in the *F. graminearum* experiment presented in publication #3. Each feature pair was assumed to be either a $[M+H]^+$, $[M+NH4]^+$, $[M+Na]^+$, and $[M+K]^+$ ion.

HRMS instruments are capable of separating different carbon-isotopologs of a compound, thus C_n can be estimated from the ratio M+1/M. This ratio is likely to be inaccurate to some extend, which may be caused by the LC-HRMS instrument itself or introduced during data analysis. Consequently, the derived number of carbon atoms from such a ratio is also inaccurate [52], which leads to putative sum formulas with an incorrect number of carbon atoms. However, if C_n can be determined exactly with the help of SIL, the number of incorrect generated sum formulas is decreased. For the examples selected, a reduction by the factor 4-27 (method "CE") compared to the same results derived for the estimated number of carbon atoms is observed. An uncertainty as little as 3% in the ratio M+1/M, which correspond to ± 3 carbon atoms, leads to 4-8 times the number of possible sum formulas and thus a high number of incorrect ones. With additional chemical knowledge in form of the Seven Golden Rules [method SGR; 24], many incorrect sum formulas can be further removed, but still a reduction by the factor 4-26 is present when the actual number of carbon atoms is known.
With the exact number of carbon atoms as determined by SIL and the seven golden rules, a total of 7,223 putative sum formulas were generated for the 704 feature pairs found in the full metabolome labelling of *F. graminearum* (publication #3). On the contrary, if the number of carbon atoms were estimated from the ratio M+1/M with a C_3 inaccuracy, 50,042 sum

formulas would have been generated, which are 7 times as many. Consequently, 42,819 (86%), of these generated sum formulas can be regarded incorrect.
Likewise, the exact number of carbon atoms also improves database searches and reduces the number of incorrect hits. This stresses that the exact number of carbon atoms, which can be determined for every feature pair with the help of ^{13}C-labelling in full metabolome labelling experiments, is very beneficial for the annotation of unknown feature pairs and metabolites.

Table 5 | Overview of sum formula generation for four randomly selected sum formulas and the results of the *F. graminearum* ^{13}C-labelling experiment presented in publication #3 with and without the exact number of carbon atoms. The number of incorrectly generated sum formulas is expressed as a factor compared to the number of putative sum formulas having the exact number of carbon atoms as determined with the help of SIL in brackets.

Sum formula	$C_{15}H_{20}O_6$		$C_{20}H_{27}N_4O_{12}P$		$C_{40}H_{66}O_5$		$C_{48}H_{80}O_{20}$		*F. graminearum*
M [u]	296.1260 u		546.1363 u		626.4910 u		976.5243 u		
Method	CE	SGR	CE	SGR	CE	SGR	CE	SGR	SGR
C_n (SIL)	1	1	14	8	2	1	31	20	7224
$C_n \pm C_3$	8 (8x)	4 (4x)	94 (7x)	55 (7x)	10 (5x)	4 (4x)	229 (7x)	156 (8x)	50,042 (7x)
$C_n \pm C_{10}$	27 (27x)	8 (8x)	301 (22x)	124 (16x)	42 (21x)	22 (22x)	715 (23x)	508 (26x)	-

In tracer-fate applications the assignment of the exact number of labelling-isotopes in a detected biotransformation product is also of great help during metabolite annotation. However, in such experiments MetExtract does not report the total number of atoms of the used labelling-element in the biotransformation product, but only the number of labelling-isotopes originating from the intact tracer or its fragment. For example, all biotransformation products of DON in wheat (publication #2) have been annotated with C_{15}. These 15 carbon atoms originated only from the studied tracer in each biotransformation product. During manual annotation several more carbon atoms were assigned to each metabolite, which originate from conjugated moieties and therefore do not contribute to the total number of labelling-isotopes. Thus, the determined number of labelling-isotopes was very helpful during manual structure interpretation of the unknown biotransformation products.

In summary, with the help of SIL novel metabolites and biotransformation products can be annotated more accurately than with just isotopolog ratios (e.g. M+1/M). In this regard, MetExtract automatically determines the number of labelling-isotopes in each detected feature pair and thus in each biotransformation product.

5.3.2. Convolution of different feature pairs into feature groups

Each detected feature pair refers to an ion species of a metabolite or one of its in-source fragments and does not necessarily represent a unique metabolite. Many metabolites, especially when they are highly abundant, are prone to result in different ion species during electrospray ionisation (e.g. $[M+H]^+$, $[M+Na]^+$). As a consequence, MetExtract automatically convolutes different ions of the same metabolite after feature pair detection into feature groups each representing a single metabolite. Since the different ion species are generated during ionisation (i.e. after chromatographic separation), their chromatographic peak shapes must be identical and only random noise may slightly distort the observed elution profiles. This fact is used to convolute the feature pairs into feature groups. If a pairwise comparison of two feature pairs reports a high correlation, these two feature pairs are very likely to originate from the same metabolite, thus they are put into the same feature group. Examples of different ion species present for the same metabolite are shown in Figures 6 and 7 in Chapter 2.

In the experiments described in the publications #2, #4, and #5, many different feature groups were manually inspected and verified. In general, this convolution was very specific and resulted in meaningful feature groups. This holds especially true for highly abundant metabolite-derived ions. For low abundant feature pairs the convolution was still specific, however, in some cases noisy chromatographic peaks were not convoluted correctly, but ended up in two separate feature groups. Also a few feature pairs, which were not baseline separated and thus showed overlapping chromatographic peak profiles with other metabolites, were not assigned to their respective feature groups. In such a case these feature pairs were assigned to separate feature groups rather than an incorrect one. However, in the evaluated and presented experiments such incorrectly grouped feature pairs were rare.

5.3.3. Feature group annotation

After feature pair convolution, commonly occurring ion species are searched for in each feature group. This search is based on the m/z difference between two feature pairs in the same feature group as well as the determined number of the labelling-isotope. The most commonly occurring ion species were $[M+H]^+$ and $[M+Na]^+$ in the positive and $[M-H]^-$ and $[M+Na-2H]^-$ in the negative ionisation mode. Any such combination allows calculating the mass of the uncharged, monoisotopic isotopolog of the intact metabolite or of its in-source fragments. In a feature group the feature with the highest mass usually represents the intact metabolite. In the optimal case, the mass of the uncharged metabolite can be determined unambiguously.

Two feature groups, which were automatically annotated with MetExtract, are depicted in Figures 6 and 7 in Chapter 2. These two metabolites show a distinct combination of different ion species and therefore allow calculating the mass of the neutral metabolite.

5.3.4. Fast polarity switching LC-HRMS data

The two ionisation polarity modes, positive and negative, result in different adducts and in-source fragments for the same metabolite. While some metabolites preferably ionise in just one of the two modes, many metabolites show different ion species in both modes. Older LC-HRMS instruments (e.g. LTQ-Orbitrap XL) are incapable of switching from one to the other ionisation polarity within a single LC-HRMS run. Thus, the same samples must be analysed sequentially in two separate LC-HRMS runs. Newer generations of ESI-devices, which are capable of fast polarity switching, support switching the ionisation polarity even within single chromatographic peaks and thus a single LC-HRMS analysis. The resulting LC-HRMS data then contains two scan-events, one for the positive and one for the negative ionisation mode and thus MS spectra derived from positive and negative ionisation polarity.
For example, the Orbitrap Exactive plus instrument used in publication #5 was configured to switch the ionisation polarity after each MS scan. With this instrument each scan was recorded in approximately half a second. Thus, each micro cycle consisting of a full MS scan in the positive and a full MS scan in the negative ionisation mode lasted for about 1 sec. This very fast acquisition times and online polarity switching capabilities of the LC-HRMS instrument enabled recording positive and negative ionisation modes quasi-parallel.

Ions derived from the same metabolite but in different ionisation polarity show perfect coelution in fast polarity switching derived LC-HRMS data. MetExtract is capable of processing such data. For this, the steps for feature pair detection are carried out separately for both ionisation modes. Successively, convolution of different feature pairs of the same metabolite is performed conjointly for the positive and negative ionisation mode results. Although there is a minimal time difference between MS scans of opposite ionisation polarity (e.g. 0.5 sec), it is still possible to compare different ions as the chromatographic peaks of metabolites are usually several times wider (e.g. 10-40 sec) than the cycle time required for a positive and negative MS scan (e.g. 1 sec). As a result, LC-HRMS data recorded with fast polarity switching may show feature groups consisting of feature pairs of both ionisation modes. Figures 6 and 7 in Chapter 2 show examples of a metabolite and a biotransformation product, which were recorded as ions of positive and negative ionisation polarity with fast polarity switching. Their respective chromatographic peaks show perfect coelution and thus MetExtract put them into the same feature groups.

Fast polarity switching is also very helpful for metabolite annotation. The different ion species observed for a compound in positive and negative ionisation polarity are complementary. Thus, fast polarity switching is useful for such metabolites, which show only one ion species in either ionisation mode. For example, in publication #5 139 metabolites were found as different ion species in either the positive (150) or negative (191) ionisation modes by MetExtract. For 19 of the 139 detected metabolites determining their uncharged, monoisotopic mass was only possible because complementary feature pairs were present in both ionisation modes.

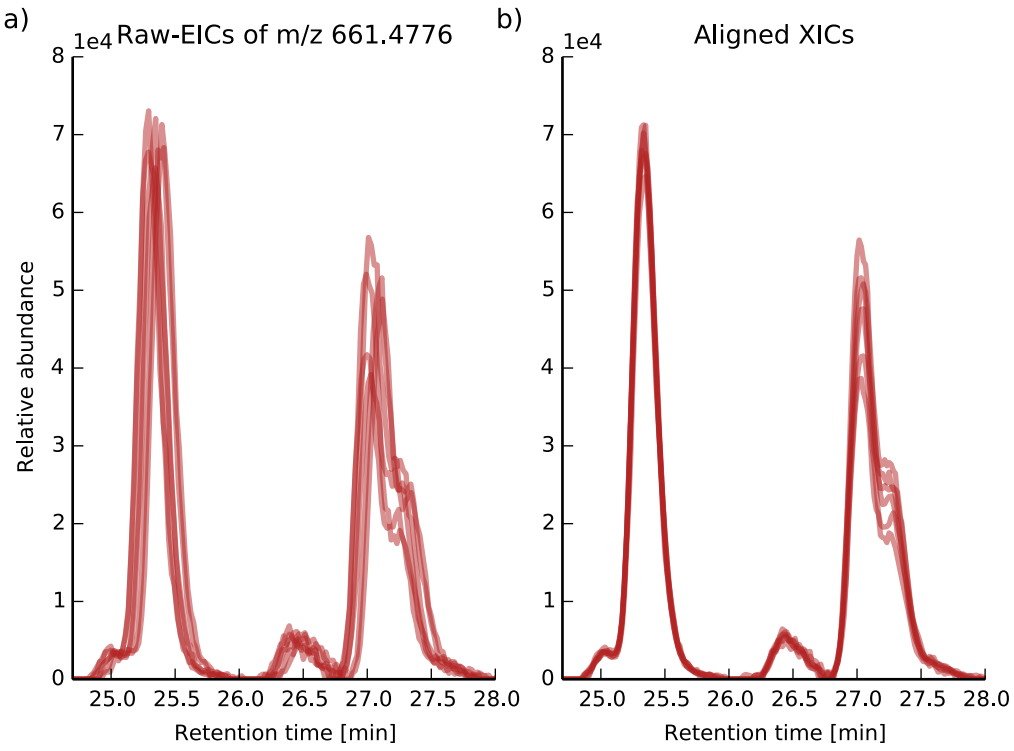


Figure 10 | a) shows the EICs of a randomly selected *m/z* value in different, biological replicates from LC-HRMS data from the *F. graminearum* experiment presented in publication #3. A slight difference in retention time is observed between the chromatographic runs, however, the shift is very little and less than the width of a typical chromatographic peak. b) shows the same EICs after chromatographic alignment with a constant time offset only (see Box 2). They show a reduced chromatographic shift.

Box 2 | Chromatographic alignment using the R-package PTW.

Chromatographic alignment: The EICs of a randomly selected metabolic feature were extracted from 6 LC-HRMS measurements from the full metabolome labelling experiment of *F. graminearum* presented in publication #3. For demonstration, these EICs were aligned with MetExtract.

5.3.5. Bracketing of sample results

For the comparison of different samples it is necessary to track the same metabolites over all samples of an experiment. The LC-HRMS data for a given experiment presented in this work was usually recorded within a single LC-HRMS sequence and Bernhard Kluger and colleagues, who were responsible for LC HRMS data acquisition, performed the analytical analysis in a controlled manner showing high repeatability (publications #2-5). Consequently, the obtained data exhibited neither a severe random shift nor a systematic shift caused by unwanted

instrument fluctuations. For example, within a LC-HRMS sequence the observed m/z accuracy was constant in all LC-HRMS data files and within the tolerated error ranges (e.g. less than 5 ppm mass deviation for LC HRMS data acquired on Orbitrap instruments). Furthermore, only a minimal retention time shift between different samples was observed, which was less than the width of a typical chromatographic peak (e.g. Figure 10a). Thus, it was not necessary to apply any normalisation steps before data processing with MetExtract. If not all samples can be measured within a sequence (i.e. measurement batch) or some kind of instrument shift occurred during a sequence, it is possible and frequently observed that the mass, retention time, or relative abundance drift severely and have to be accounted for ahead of data processing with MetExtract [103].

With respect to chromatographic shifts, MetExtract includes an optional alignment method to account for putative shifts of retention times throughout the analysed LC-HRMS data. For this alignment step, the R-package PTW [polynomial time warping; 60] is utilised. This alignment method uses a polynomial function to adjust different chromatograms so they correlate with a reference chromatogram. Depending the polynoms' order, PTW even allows accounting for non-constant or non-linear chromatographic shifts. After applying the PTW algorithm the aligned chromatograms show reduced retention time diversity. For example, 6 unaligned EICs of randomly selected feature pairs from the *F. graminearum* experiment (publication #3) are shown in Figure 10a. A slight retention time shift is obvious. This observed shift, however, is less than the width of a typical chromatographic peak and needs no correction. The same EICs were then aligned as described in Box 2 and are shown in Figure 10b. These aligned EICs show a reduced chromatographic shift and nearly perfect coelution in all samples. This simple example illustrates that retention time shifts can be efficiently handled in MetExtract.

In summary, MetExtract automatically brackets different feature pairs from all acquired LC-HRMS data files within an experiment. Furthermore, the software allows accounting for different forms of retention time shifts using the PTW package and automatically integrates feature pairs not detected initially by strict data processing settings.

5.4. Improved relative quantification for comparison of different experimental conditions

In LC-ESI-HRMS analysis, coeluting substances can either decrease or enhance the ionisation efficiency of other compounds. Such effects are mainly caused by the fact that ESI can only ionise a limited amount of molecules at a time [49, 104]. Consequently, different metabolome compositions of different samples and experimental conditions lead to ionisation enhancement or suppression effects. Since these manifest themselves in the ionisation source, matrix effects can be very diverse across an entire chromatogram. As a consequence, the observed metabolite abundances in different biological conditions are distorted to a certain extend. This is especially problematic, when the aim of the experiment is to compare metabolite

abundances in different conditions. For example, coeluting substances $X_{1..I}$ in condition A may alter the ionisation efficiency of a certain metabolite X_j. In another condition B, the relative amount of the metabolites $X_{1..i}$ may be different than in A and thus the ionisation efficiency of X_j is decreased in B, although the absolute concentration of X_j is equal in both conditions A and B. Additionally, the samples of condition B may contain some metabolites $X_{i+1..i+k}$, which coelute with metabolite X_j and further decrease its ionisation efficiency. Subsequent statistical analysis of X_j could report a statistically significant abundance difference between the conditions A and B. However, in this constructed example, this difference is not of biological origin but rather a result of matrix effects caused by coeluting metabolites.
Moreover, when a metabolomics experiment cannot be analysed within a single LC-HRMS sequence, the recorded, relative metabolite abundances may also be distorted across the different LC-HRMS sequences. Especially after instrument recalibration, the observed, relative metabolite abundances can be quite different in comparison to previously analysed biological conditions.
As a result, a difference of a metabolite between two or more conditions cannot be directly referred to a biological origin. Before a sound statistical investigation can be performed the magnitude of different matrix effects and putative instrument shifts must be investigated and accounted for.

In this respect, SIL is considered the gold standard to resolve different ionisation effects. In targeted approaches it allows determining the absolute concentration of the few studied metabolites; in untargeted metabolomics experiments a metabolome- and experiment-wide internal standardisation is well suited for improving relative quantification across different samples or experimental states. In the developed full metabolome labelling workflow this internal standardisation is achieved by spiking an equal amount of the pooled, labelled sample or the labelled reference material to each studied, native sample during sample preparation (Figure 3 in Chapter 2). During LC-HRMS analysis, different matrix effects affect both the native and the labelled metabolites equally, while their ratios are unaffected. As a result, SIL assisted internal standardisation effectively accounts for different matrix effects of each detected metabolite. Although this greatly improves the comparison of different samples, it does not allow calculating the absolute concentration of the respective metabolites.
To facilitate relative metabolite quantification with the help of metabolome- and experiment-wide internal standardisation, MetExtract automatically integrates the chromatographic peaks of native and labelled ions of feature pairs and determines their fold ratio, which represents matrix effect free, relative metabolite abundances.

An example, in which metabolome- and experiment-wide internal standardisation of different *F. graminearum* genotypes helped to improve multivariate, statistical analysis using PCA is presented in publication #3. There, the clustering of aggregate samples was improved when PCA was calculated using the relative metabolite abundances after metabolome- and experiment-wide internal standardisation.

Box 3 | Continued ^{13}C-full metabolome labelling experiment of wheat as presented in publication #3. Details are given to a degree necessary to understand the performed experiment with respect to the presented results.

Biological experiment: Ears of the wheat genotypes CM38 and CM51 were treated with the pathogenic fungi *F. graminearum* (substring "Fus" in the sample name), with water (substring "Moc"), and no treatment (substring "Non") at time point zero with 3 biological replicates each. Samples were taken 96 hours after treatment.
Sample preparation and LC-HRMS analysis: Each sample was standardised internally with commercially acquired ^{13}C-labelled wheat material from IsoLife (Wageningen, The Netherlands; http://www.isolife.eu; last accessed 1st September 2014). Sample preparation and LC-HRMS measurements were performed as described in Bueschl [3]. The 18 samples were randomly assigned to three independent LC-HRMS sequences and measured on an LTQ-Orbitrap XL instrument. In the first measurement sequence 7 samples were measured (red samples). Then, the instrument was used otherwise for 4 weeks, during which it was shut down and recalibrated. After that 6 samples were analysed in the second LC-HRMS sequence (green samples). The remaining 5 samples were subsequently measured in a third LC-HRMS sequence immediately after the second sequence (blue samples).
Data processing: Acquired LC-HRMS data were processed with MetExtract as described in publication #3 for the wheat samples.
Multivariate statistics: HCA (range-scaling; squared Euclidean distance and ward linkage) was calculated with the relative abundances of the native metabolite ions (Figure 11a). Additionally, HCA was performed with the internal standardisation corrected feature pair ratios using the same metabolites as for the first HCA (Figure 11b).

In an untargeted wheat experiment (continuation from the wheat experiment presented in publication #3; manuscript in preparation), in which samples consisting of native and globally ^{13}C-labelled wheat extracts were measured in different LC-HRMS sequences (Box 3), metabolome- and experiment-wide internal standardisation also efficiently corrected relative metabolite abundances and consequently improved statistical investigation. The samples analysed in measurement sequence 1 showed different metabolite abundances than those samples analysed in measurement sequences 2 and 3. This comparison of the different experimental conditions is depicted in form of a HCA dendrogram. Based on the raw, monoisotopic feature intensities, the samples from the first sequence form a cluster and the samples from the second and third sequences form another one (Figure 11a). This is caused by a change in the detector sensitivity between sequence 1 and sequences 2 and 3. After internal standardisation of the detected metabolites, the sequence specific separation of the analysed samples is accounted for. In this second dendrogram (Figure 11b) all *F. graminearum* treated samples show a clear separation from the water and non-treated samples and no sequence specific separations of the samples are observable.

To summarise, metabolome- and experiment-wide internal standardisation using either pooled and labelled samples or labelled reference material provides a convenient way to account for different matrix effects and MS detector alterations observed for different measurement sequences. In this respect, MetExtract automatically integrates the necessary chromatographic peaks and determines internal standardisation corrected fold ratios, which greatly improve successive statistical comparison.

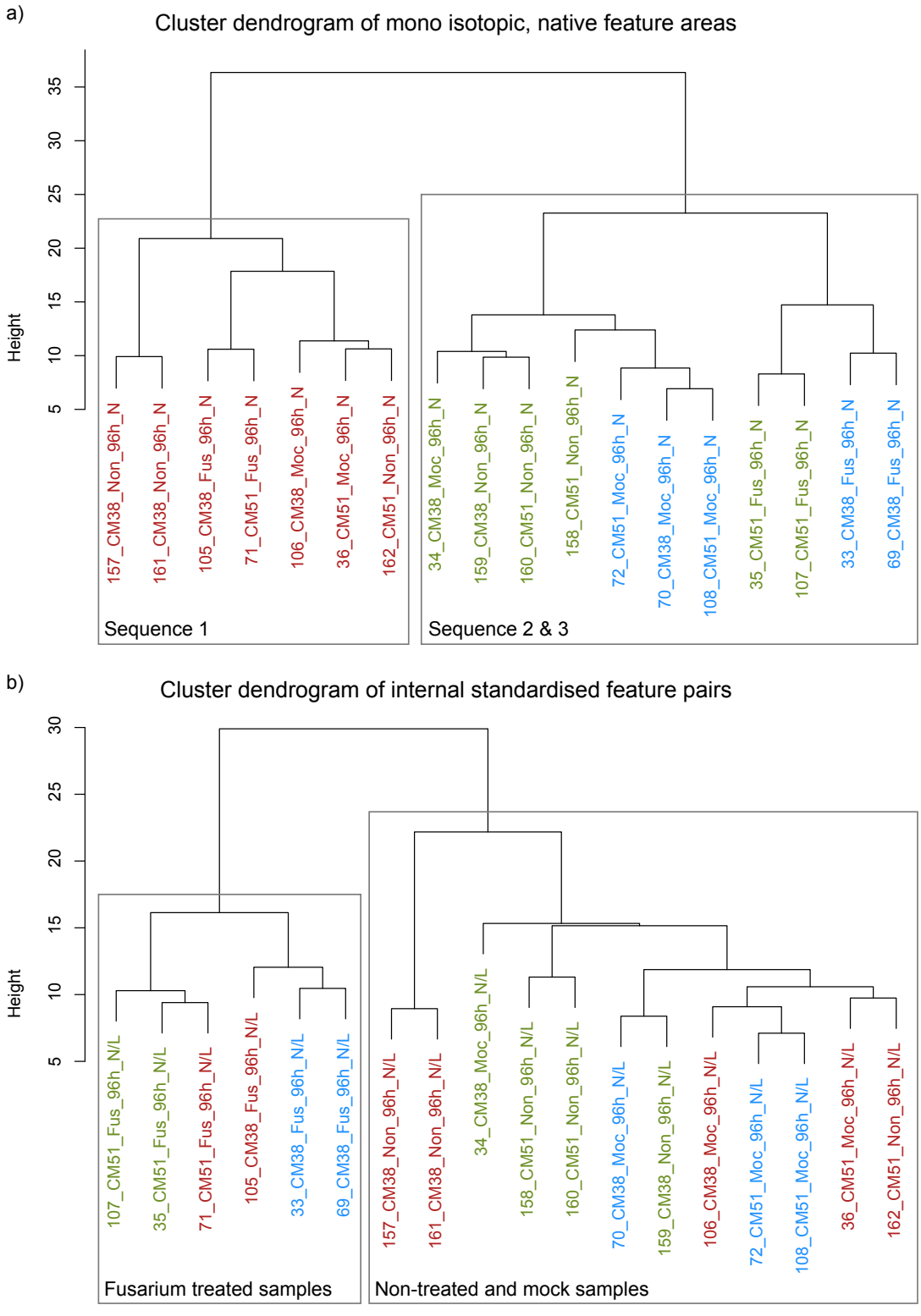


Figure 11 | HCA dendrograms before (a) and after (b) SIL assisted internal standardisation. While the analysed samples in a) cluster according to their measurement sequence, in b) the *F. graminearum* treated samples are separated from the mock and non-treated samples and a sequence specific clustering is not observed (see Box 3 for experimental details).

5.5. Detection of tracer-derived biotransformation products

SIL is perfectly suited for studying the metabolic fate of tracer substances. A biological system, which has access to native and labelled tracer substances, metabolises both forms [68]. As a result, all tracer-derived biotransformation products are also present in their native and partly labelled forms, while metabolites that are not derived from the studied tracer substance are only present in their native forms. This fact enables the efficient search for all tracer-derived biotransformation products.

The isotope pattern of native and labelled biotransformation products observed with LC-HRMS data is in certain aspects different from that of a globally labelled metabolite in full metabolome labelling experiments: Since only the tracer-derived part of a biotransformation product can be labelled, it is not possible to derive the total number of atoms of the used labelling-element in a respective metabolic feature. Consequently, the isotope pattern of such labelled metabolites contains parts originating from the labelled tracer and native parts from conjugated moieties. An example of such a biotransformation product in depicted in Figure 7 in Chapter 2. It shows a detoxification product of DON in wheat (publication #2).

For the detection of tracer-derived biotransformation products, only the recognition of isotope patterns and their isotopolog abundance verification (step 1) were required to be adapted compared to the detection of globally labelled metabolite-derived ions in full metabolome labelling experiments. Chromatographic separation of different biotransformation product homologs and convolution of different feature pairs from the same metabolite into feature groups remained unchanged, as these steps only use the monoisotopic and the uniformly/consistently labelled isotopologs of the same metabolite.

Additionally, an optional M:M' ratio check was implemented, as the M:M' ratio of biotransformation products derived from an exogenous tracer must approximately match the M:M' ratio of the applied, pure tracer.

This extended version of MetExtract for tracer-fate studies was used to analyse the metabolic fate of the exogenous mycotoxin HT2 in barley plants (manuscript in preparation; Box 4). A total of 9 different metabolites each containing either an intact HT2 molecule or its fragments were detected (Figure 12). Furthermore, the relative ratios of native and ^{13}C-labelled biotransformation products (M:M' ratio) are approximately equal for each metabolite (i.e. same relative abundance of the respective chromatographic peaks) and correspond well to the applied HT2 ratio (1:1/v:v).

In summary SIL assisted tracer-fate studies allow detecting known and unknown biotransformation products of the studied tracer in virtually any biological system. In this regard, MetExtract automatically detects such biotransformation products, as these are the only metabolites present in both native and labelled forms. All other metabolites, which are not descendent of the respective tracer, are efficiently filtered.

Box 4 | Biological experiment for studying the metabolic fate of the mycotoxin HT2 in barley. Details are given to a degree necessary to understand the performed experiment with respect to the presented results.

Biological experiment: Both native and globally ^{13}C-labelled HT2 were applied (1:1/v:v) to ears of flowering barley plants, which metabolise and detoxify this mycotoxin (similar to publicaiton #2).
LC-HRMS analysis: After sample preparation the barley extracts containing the biotransformation products of native and globally ^{13}C-labelled HT2 were analysed with a Quadrupole-Time-of-Flight instrument (Agilent Technologies™) operated in positive ionisation mode.
Data processing: Acquired LC-HRMS data files were processed with MetExtract. Settings are provided in Table 4 row T2 / HT2 in wheat / barley.

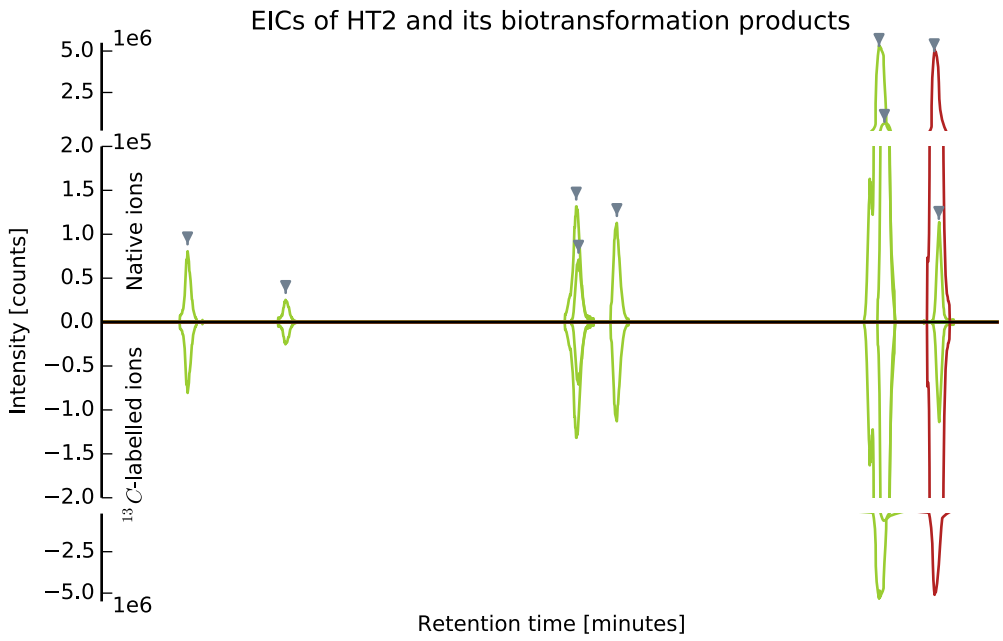


Figure 12 | EICs of native, monoisotopic (positive intensity) and consistently, ^{13}C-labelled (negative intensities) HT2-derived biotransformation products in barley. 8 such biotransformation products (green) were detected besides the unprocessed HT2 (red). The illustration's ordinate is interrupted twice for improved illustration of the differently abundant biotransformation products.

6. Further development and current applications of MetExtract

In the context of this thesis, MetExtract a novel software tool for the analysis of LC-HRMS data acquired for SIL assisted untargeted metabolomics experiments was implemented and utilised. It automatically processes LC-HRMS data from full metabolome labelling and tracer-fate experiments involving the concurrent measurement of native and labelled sample material.

The current version of MetExtract only supports such native and labelled metabolites and biotransformation products, which have clearly separated isotope patterns of the native and labelled metabolites. Moreover, it requires a high isotopic enrichment of the labelling-isotope. For lower isotopic enrichment (less than 97%), isotopic patterns start to overlap for certain metabolites. Additionally, the fully labelled isotopolog of a metabolite-derived ion cannot easily be recognised, which distorts the annotation with the exact number of labelled isotopes. To support this type of data in upcoming releases, MetExtract will be adapted accordingly. Other examples of planned features for further releases include support for i) concurrent labelling with two different isotopes (e.g. ^{13}C and ^{15}N), ii) the concurrent study of several tracer substances in the same experiment, and iii) the deconvolution of overlapping isotope patterns from different elements.

With respect to biological applications, the developed workflows and MetExtract are currently being used intensively in the SFB project "Fusarium" for studying the LC-HRMS accessible metabolomes of wheat, maize, and *F. graminearum* as well as the metabolic interaction of *F. graminearum* and the plants (manuscript in preparation; [105]). To this end, preliminary experiments for inhouse cultivation of ^{13}C-labelled wheat material were performed, which will allow infecting both non-labelled and ^{13}C-labelled wheat plants with the pathogen in the future. Furthermore, the presented approaches will be used to study fundamental aspects of analytical workflow optimisation.
Besides these studies, the software will further be used for investigating the metabolic fate of the mycotoxins T2/HT2 in wheat and barley, for which experiments have already been performed successfully. Additionally, biological interpretation of the phenylalanine-derived metabolites in wheat cell suspension cultures will be done. Additional experiments similar to those presented in publication #5 will be carried out including a larger number of biological replicates to allow reliable and sound statistical investigation.

7. Conclusion

In recent years, SIL became increasingly used for many applications in untargeted metabolomics research. Among others its advantages include the global and efficient detection of only truly biology-derived metabolites, annotation of unknown metabolic features with the number of atoms of the used labelling-isotope, and experiment- and metabolome-wide internal standardisation for improved comparative quantification. Furthermore, SIL is perfectly suited for studying the metabolic fate of virtually any tracer-substance in a biological system.

Software tools for the automatic processing of SIL derived metabolomics data are still rare. Therefore, novel SIL assisted workflows and software tools for the processing of the acquired LC-HRMS data were developed and used. These works were carried out in an interdisciplinary and experiment-oriented cooperation between analytical chemists and bioinformaticians.

In this regards, the presented doctoral thesis details a novel software tool, named MetExtract, which is designed for the automated evaluation of SIL-derived LC-HRMS data from either full metabolome labelling experiments or tracer-fate studies of exogenous or endogenous compounds. It requires LC-HRMS data from samples consisting of both native and labelled metabolites and supports different labelling-isotopes (e.g. ^{13}C, ^{15}N, ^{34}S) with a degree of isotopic enrichment of 97% or higher. MetExtract uses the open-data format mzXML to read raw-LC-HRMS data and thus works with virtually any LC-HRMS platform. Moreover, the software is capable of processing fast polarity switching derived LC-HRMS data.

MetExtract detects metabolites using their unique isotope patterns of native and labelled metabolite-derived ions. Detected feature pairs are intensively verified with their respective isotopologs and a check for perfect chromatographic coelution. These validations of the SIL characteristics ensure that only truly biological metabolites are extracted while efficiently discarding any unspecific signals or signals originating from contaminants. Furthermore, MetExtract is highly specific and only reports a very low number of incorrectly detected feature pairs even in samples with a high background. As a result, SIL-assisted, untargeted metabolomics in combination with MetExtract for the first time enables the holistic detection of only truly metabolite-derived ions, which constitutes a great step towards the unbiased annotation of all via LC-HRMS accessible metabolites of the biological system under investigation.

In tracer-fate studies, MetExtract detects only biotransformation products derived from the native and labelled tracer. The detected biotransformation products are annotated with the number of atoms of that part, which is derived from the studied tracer. This annotation notably enhances subsequently performed manual structural elucidation.

After detection, different metabolic features derived from the same metabolite are automatically convoluted into feature groups facilitating a holistic annotation of detected metabolites. These feature groups are annotated with commonly observed ion species and putative neutral losses of in-source fragments are calculated using the determined number of labelling-isotopes in each metabolic feature. Consequently, this also reduces the number of incorrectly annotated

database hits or possible sum formulas of unknown metabolites and thus improves putative metabolite identification and characterisation.

Besides its use for the efficient detection of metabolites, SIL is also regarded the gold standard to account for different matrix effects. With appropriate sample preparation and pooling protocols ahead of LC-HRMS analysis, the software automatically integrates all detected native and corresponding labelled features. The fold ratios of these two feature areas are subsequently used to correct for different ionisation enhancement or suppression effects of the mass spectrometer. Moreover, global internal standardisation also enables the compensation of analytical shifts or drifts within and across different measurement sequences. This internal standardisation is especially helpful in subsequent statistical analysis and biological interpretation of the respective untargeted metabolomics experiments and can be accessed in an automated fashion using MetExtract.

The benefits of SIL-assisted, untargeted metabolomics in combination with MetExtract have been demonstrated with *F. graminearum*, wheat and maize experiments. Of the metabolites detected by MetExtract, only a small number (less than five per sample) represent incorrectly detected feature pairs. Most of these false positives originated from Fourier transform artefacts, which are a common source of bias in LC HRMS data. Additionally, application in the *F. graminearum* study showed that internal standardisation with pooled material from all labelled fungi samples improved analytical precision and multivariate statistical investigation.

Furthermore, MetExtract was used for detecting known and unknown biotransformation products derived from different tracer substances. In a first experiment, which studied the detoxification mechanism of wheat plants stressed with the mycotoxin DON, in total 9 detoxification products were found, 5 of which have not been reported previously.

In a similar experiment, which studied the metabolic fate of phenylalanine in grape-berries, 63 phenylalanine-containing metabolites were automatically detected, 13 of which showed either increased or decreased abundances after the grape-berries were exposed to heat.

Furthermore, the metabolic fate of phenylalanine was also studied in wheat-cell suspension cultures. In these samples 341 phenylalanine-containing metabolic features, which were derived from fast polarity switching ionisation and automatically convoluted to 139 different metabolites, were extracted. Several of these metabolites showed either increased or decreased abundance after the cell suspension cultures have been stressed with DON. Additionally, in this experiment it has been shown that fast polarity switching is highly beneficial for metabolite detection and subsequent annotation.

In conclusion, SIL-assisted, untargeted metabolomics research has many advantages over conventional, labelling-free methods. With MetExtract these benefits can be accessed in a structured and automated fashion. Together with the presented workflows for full metabolome labelling and tracer-fate studies, MetExtract constitute a major step forward for SIL-assisted, untargeted metabolomics research, which is gaining popularity in many metabolomics applications.

References

1. Bueschl C, Krska R, Kluger B, Schuhmacher R: **Isotopic labeling-assisted metabolomics using LC-MS**. *Analytical and Bioanalytical Chemistry* 2013, **405**(1): 27-33.

2. Kluger B, Bueschl C, Lemmens M, Berthiller F, Haubl G, Jaunecker G, Adam G, Krska R, Schuhmacher R: **Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat**. *Analytical and Bioanalytical Chemistry* 2013, **405**(15): 5031-5036.

3. Bueschl C, Kluger B, Lemmens M, Adam G, Wiesenberger G, Maschietto V, Marocco A, Strauss J, Bodi S, Thallinger GG, Krska R, Schuhmacher R: **A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research**. *Metabolomics* 2014, **10**(4): 754-769.

4. Chassy AW, Bueschl C, Lee H, Lerno L, Oberholster A, Barile D, Schuhmacher R, Waterhouse AL: **Tracing flavonoid degradation in grapes by MS filtering with stable isotopes**. *Food Chemistry* 2015, **166**: 448-455.

5. Kluger B, Bueschl C, Neumann N, Stuckler R, Doppler M, Chassy AW, Waterhouse AL, Rechthaler J, Kampleitner N, Thallinger GG, Adam G, Krska R, Schuhmacher R: **Untargeted Profiling of Tracer-Derived Metabolites Using Stable Isotopic Labeling and Fast Polarity-Switching LC-ESI-HRMS**. *Analytical Chemistry* 2014, **86**(23): 11533-11537.

6. Weckwerth W: **Metabolomics in systems biology**. *Annual Review of Plant Biology* 2003, **54**: 669-689.

7. Oliver SG, Winson MK, Kell DB, Baganz F: **Systematic functional analysis of the yeast genome**. *Trends Biotechnol* 1998, **16**(9): 373-378.

8. Fiehn O: **Metabolomics – the link between genotypes and phenotypes**. *Plant Molecular Biology* 2002, **48**(1): 155-171.

9. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW: **Potential of metabolomics as a functional genomics tool**. *Trends in Plant Science* 2004, **9**(9): 418-425.

10. Cochereau D, Junot C: **Metabolomics contribution to predictive biomarker discovery**. *Oncologie* 2013, **15**(9): 461-466.

11. Beecher C: **Metabolomics techniques for biomarker discovery**. *Abstracts of Papers of the American Chemical Society* 2011, **242**.

12. Terashima Y, Nishiumi S, Minami A, Kawano Y, Hoshi N, Azuma T, Yoshida M: **Metabolomics-based search for therapeutic agents for non-alcoholic steatohepatitis**. *Archives of Biochemistry and Biophysics* 2014, **555-556**: 55-65.

13. Sakai A, Nishiumi S, Shiomi Y, Kobayashi T, Izumi Y, Kutsumi H, Hayakumo T, Azuma T, Yoshida M: **Metabolomic analysis to discover candidate therapeutic agents against acute pancreatitis**. *Archives of Biochemistry and Biophysics* 2012, **522**(2): 107-120.

14. Vaidyanathan S, Harrigan GG, Goodacre R: **Metabolome analyses : strategies for systems biology**; 2005: 390 p.

15. Robertson DG, Frevert U: **Metabolomics in drug discovery and development**. *Clinical Pharmacology and Therapeutics* 2013, **94**(5): 559-561.

16. Watkins SM, German JB: **Metabolomics and biochemical profiling in drug discovery and development**. *Current Opinion in Molecular Therapeutics* 2002, **4**(3): 224-228.

17. Di Lena M, Travaglio E, Altomare DF: **Metabolomics: a potential powerful ally in the fight against cancer**. *Colorectal Disease* 2014, **16**(4): 235-238.

18. D'alessandro A, Zolla L: **Proteomics and metabolomics in cancer drug development**. *Expert Review of Proteomics* 2013, **10**(5): 473-488.

19. Fan TWM, Lane AN, Higashi RM: **The promise of metabolomics in cancer molecular therapeutics**. *Current Opinion in Molecular Therapeutics* 2004, **6**(6): 584-592.

20. Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG: **A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations**. *Nature Biotechnology* 2001, **19**(1): 45-50.

21. Robertson DG, Watkins PB, Reily MD: **Metabolomics in Toxicology: Preclinical and Clinical Applications**. *Toxicological Sciences* 2011, **120**: S146-S170.

22. Nicholson JK, Lindon JC, Holmes E: **'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data**. *Xenobiotica* 1999, **29**(11): 1181-1189.

23. Warth B, Sulyok M, Fruhmann P, Berthiller F, Schuhmacher R, Hametner C, Adam G, Frohlich J, Krska R: **Assessment of human deoxynivalenol exposure using an LC-MS/MS based biomarker method**. *Toxicology Letters* 2012, **211**(1): 85-90.

24. Kind T, Fiehn O: **Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry**. *BMC Bioinformatics* 2007, **8**: 105-124.

25. Milne SB, Mathews TP, Myers DS, Ivanova PT, Brown HA: **Sum of the Parts: Mass Spectrometry-Based Metabolomics**. *Biochemistry* 2013, **52**(22): 3829-3840.

26. Dunn WB, Ellis DI: **Metabolomics: Current analytical platforms and methodologies**. *Trends in Analytical Chemistry* 2005, **24**(4): 285-294.

27. Koal T, Deigner HP: **Challenges in Mass Spectrometry Based Targeted Metabolomics**. *Current Molecular Medicine* 2010, **10**(2): 216-226.

28. Trethewey RN: **Metabolite profiling as an aid to metabolic engineering in plants**. *Current Opinion in Plant Biology* 2004, **7**(2): 196-201.

29. Kossel A: **Ueber die chemische Zusammensetzung der Zelle**. *Archiv für Physiologie*; 1891: 181-186.

30. Fraenkel GS: **The raison d'etre of secondary plant substances; these odd chemicals arose as a means of protecting plants from insects and now guide insects to food**. *Science* 1959, **129**(3361): 1466-1470.

31. Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics**. *Mass Spectrometry Reviews* 2007, **26**(1): 51-78.

32. Smilde AK, van der Werf MJ, Bijlsma S, van der Werff-van-der Vat BJC, Jellema RH: **Fusion of mass spectrometry-based metabolomics data**. *Analytical Chemistry* 2005, **77**(20): 6729-6736.

33. Theodoridis GA, Gika HG, Want EJ, Wilson ID: **Liquid chromatography–mass spectrometry based global metabolite profiling: A review**. *Analytica Chimica Acta* 2012, **711**(0): 7-16.

34. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G: **XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification**. *Analytical Chemistry* 2006, **78**(3): 779-787.

35. Tautenhahn R, Bottcher C, Neumann S: **Highly sensitive feature detection for high resolution LC/MS**. *BMC Bioinformatics* 2008, **9**: 504-519.

36. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G: **XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data**. *Analytical Chemistry* 2012, **84**(11): 5035-5039.

37. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S: **CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets**. *Analytical Chemistry* 2012, **84**(1): 283-289.

38. Katajamaa M, Miettinen J, Oresic M: **MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data**. *Bioinformatics* 2006, **22**(5): 634-636.

39. Pluskal T, Castillo S, Villar-Briones A, Oresic M: **MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data**. *BMC Bioinformatics* 2010, **11**: 395-404.

40. Bertsch A, Gropl C, Reinert K, Kohlbacher O: **OpenMS and TOPP: open source software for LC-MS data analysis**. *Methods in Molecular Biology* 2011, **696**: 353-367.

41. Sturm M, Bertsch A, Gropl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-Trieglaff O, Zerck A, Reinert K, Kohlbacher O: **OpenMS - an open-source software framework for mass spectrometry**. *BMC Bioinformatics* 2008, **9**: 163-173.

42. Lommen A: **MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing**. *Analytical Chemistry* 2009, **81**(8): 3079-3086.

43. Lommen A, Kools HJ: **MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware**. *Metabolomics* 2012, **8**(4): 719-726.

44. Hartler J, Trotzmuller M, Chitraju C, Spener F, Kofeler HC, Thallinger GG: **Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data**. *Bioinformatics* 2011, **27**(4): 572-577.

45. Clasquin MF, Melamud E, Rabinowitz JD: **LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine**. *Current Protocols in Bioinformatics* 2012, **14**(11): 1-23.

46. Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R: **PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis**. *Analytical Chemistry* 2011, **83**(7): 2786-2793.

47. Keller BO, Suj J, Young AB, Whittal RM: **Interferences and contaminants encountered in modern mass spectrometry**. *Analytica Chimica Acta* 2008, **627**(1): 71-81.

48. Böttcher C, Roepenack-Lahaye E, Willscher E, Scheel D, Clemens S: **Evaluation of Matrix Effects in Metabolite Profiling Based on Capillary Liquid Chromatography Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry**. *Analytical Chemistry* 2007, **79**(4): 1507-1513.

49. Taylor PJ: **Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry**. *Clinical Biochemistry* 2005, **38**(4): 328-334.

50. Annesley TM: **Ion Suppression in Mass Spectrometry**. *Clinical Chemistry* 2003, **49**(7): 1041-1044.

51. Krauss M, Singer H, Hollender J: **LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns**. *Analytical and Bioanalytical Chemistry* 2010, **397**(3): 943-951.

52. Xu Y, Heilier JF, Madalinski G, Genin E, Ezan E, Tabet JC, Junot C: **Evaluation of Accurate Mass and Relative Isotopic Abundance Measurements in the LTQ-Orbitrap Mass Spectrometer for Further Metabolomics Database Building**. *Analytical Chemistry* 2010, **82**(13): 5490-5501.

53. Liland KH: **Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis**. *Trends in Analytical Chemistry* 2011, **30**(6): 827-841.

54. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD: **Analytical and statistical approaches to metabolomics research**. *Journal of Separation Science* 2009, **32**(13): 2183-2199.

55. Kruve A, Leito I: **Comparison of different methods aiming to account for/overcome matrix effects in LC/ESI/MS on the example of pesticide analyses**. *Analytical Methods* 2013, **5**(12): 3035-3044.

56. Broadhurst DI, Kell DB: **Statistical strategies for avoiding false discoveries in metabolomics and related experiments**. *Metabolomics* 2006, **2**(4): 171-196.

57. Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Calmani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjostrom M, Trygg J, Wulfert F: **Proposed minimum reporting standards for data analysis in metabolomics**. *Metabolomics* 2007, **3**(3): 231-241.

58. Madsen R, Lundstedt T, Trygg J: **Chemometrics in metabolomics-A review in human disease diagnosis**. *Analytica Chimica Acta* 2010, **659**(1-2): 23-33.

59. van den Berg R, Hoefsloot H, Westerhuis J, Smilde A, van der Werf MJ: **Centering, scaling, and transformations: improving the biological information content of metabolomics data**. *BMC Genomics* 2006, **7**(1): 142-156.

60. Wehrens R: **Chemometrics With R: Multivariate Data Analysis in the Natural Sciences and Life Sciences**. *Use R!*; 2011.

61. Baillie TA: **The Use of Stable Isotopes in Pharmacological Research**. *Pharmacological Reviews* 1981, **33**(2): 81-132.

62. Janghorbani M, Weaver CM, Ting BTG, Young VR: **Labeling of Soybeans with the Stable Isotope Zn-70 for Use in Human Metabolic Studies**. *Journal of Nutrition* 1983, **113**(5): 973-978.

63. Wu YL, Gretz MR: **Stable-Isotope Labeling Method for Studies of Saccharide Metabolism in *Agardhiella-Subulata***. *Hydrobiologia* 1993, **261**: 595-600.

64. Strong JM, Upton DK, Anderson LW, Monks A, Chisena CA, Cysyk RL: **A Novel-Approach to the Analysis of Mass Spectrally Assayed Stable Isotope-Labeling Experiments**. *Journal of Biological Chemistry* 1985, **260**(7): 4276-4281.

65. Iglesias J, Sleno L, Volmer DA: **Isotopic Labeling of Metabolites in Drug Discovery Applications**. *Current Drug Metabolism* 2012, **13**(9): 1213-1225.

66. Hiller K, Metallo C, Stephanopoulos G: **Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics**. *Current Pharmaceutical Biotechnology* 2011, **12**(7): 1075-1086.

67. Winter G, Kromer JO: **Fluxomics - connecting 'omics analysis and phenotypes**. *Environ Microbiol* 2013, **15**(7): 1901-1916.

68. Klein S, Heinzle E: **Isotope labeling experiments in metabolomics and fluxomics**. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine* 2012, **4**(3): 261-272.

69. Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L: **High-Resolution Direct Infusion-Based Mass Spectrometry in Combination with Whole ^{13}C Metabolome Isotope Labeling Allows Unambiguous Assignment of Chemical Sum Formulas**. *Analytical Chemistry* 2008, **80**(24): 9417-9425.

70. Giavalisco P, Li Y, Matthes A, Eckhardt A, Hubberten H, Hesse H, Segu S, Hummel J, Köhl K, Willmitzer L: **Elemental formula annotation of polar and lipophilic metabolites using ^{13}C, ^{15}N and ^{34}S isotope labelling, in combination with high-resolution mass spectrometry**. *The Plant Journal* 2011, **68**(2): 364-376.

71. Cano PM, Jamin EL, Tadrist S, Bourdaud'hui P, Pean M, Debrauwer L, Oswald IP, Delaforge M, Puel O: **New untargeted metabolic profiling combining mass spectrometry and isotopic labeling: application on *Aspergillus fumigatus* grown on wheat**. *Analytical Chemistry* 2013, **85**(17): 8412-8420.

72. Giavalisco P, Koehl K, Hummel J, Seiwert B, Willmitzer L: **^{13}C Isotope-Labeled Metabolomes Allowing for Improved Compound Annotation and Relative**

Quantification in Liquid Chromatography-Mass Spectrometry-based Metabolomic Research. *Analytical Chemistry* 2009, **81**(15): 6546-6551.

73. Baillie TA, Rettenmeier AW: **Recent Advances in the Use of Stable Isotopes in Drug-Metabolism Research**. *Journal of Clinical Pharmacology* 1986, **26**(6): 481-484.

74. Baty JD, Lindsay RM, Fox WR, Willis RG: **Stable Isotopes as Probes for the Metabolism of Acetanilide in Man and the Rat**. *Biomedical and Environmental Mass Spectrometry* 1988, **16**(1-12): 183-189.

75. Fan TWM, Lorkiewicz PK, Sellers K, Moseley HNB, Higashi RM, Lane AN: **Stable isotope-resolved metabolomics and applications for drug development**. *Pharmacology & Therapeutics* 2012, **133**(3): 366-391.

76. Anderson RA, Baillie TA, Axelson M, Cronholm T, Sjovall K, Sjovall J: **Stable Isotope Studies on Steroid-Metabolism and Kinetics - Sulfates of 3-Alpha-Hydroxy-5-Alpha-Pregnane Derivatives in Human-Pregnancy**. *Steroids* 1990, **55**(10): 443-457.

77. Li F, Pang XY, Krausz KW, Jiang CT, Chen C, Cook JA, Krishna MC, Mitchell JB, Gonzalez FJ, Patterson AD: **Stable Isotope- and Mass Spectrometry-based Metabolomics as Tools in Drug Metabolism: A Study Expanding Tempol Pharmacology**. *Journal of Proteome Research* 2013, **12**(3): 1369-1376.

78. Hiller K, Wegner A, Weindl D, Cordes T, Metallo CM, Kelleher JK, Stephanopoulos G: **NTFD-a stand-alone application for the non-targeted detection of stable isotope-labeled compounds in GC/MS data**. *Bioinformatics* 2013, **29**(9): 1226-1228.

79. Hiller K, Metallo CM, Kelleher JK, Stephanopoulos G: **Nontargeted Elucidation of Metabolic Pathways Using Stable-Isotope Tracers and Mass Spectrometry**. *Analytical Chemistry* 2010, **82**(15): 6621-6628.

80. Huang X, Chen YJ, Cho K, Nikolskiy I, Crawford PA, Patti GJ: **X^{13}CMS: global tracking of isotopic labels in untargeted metabolomics**. *Analytical Chemistry* 2014, **86**(3): 1632-1639.

81. Chokkathukalam A, Jankevics A, Creek DJ, Achcar F, Barrett MP, Breitling R: **mzMatch–ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data**. *Bioinformatics* 2013, **29**(2): 281-283.

82. Guo K, Li L: **Differential ^{12}C-/^{13}C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome**. *Analytical Chemistry* 2009, **81**(10): 3919-3932.

83. Rosenblum C: **Principles of Isotope Dilution Assays**. *Analytical Chemistry* 1957, **29**(12): 1740-1744.

84. Varga E, Glauner T, Koppen R, Mayer K, Sulyok M, Schuhmacher R, Krska R, Berthiller F: **Stable isotope dilution assay for the accurate determination of mycotoxins in maize by UHPLC-MS/MS**. *Analytical and Bioanalytical Chemistry* 2012, **402**(9): 2675-2686.

85. Neumann NKN, Lehner SM, Kluger B, Bueschl C, Sedelmaier K, Lemmens M, Krska R, Schuhmacher R: **Automated LC-HRMS(/MS) approach for the annotation of fragment ions derived from stable isotope labeling-assisted untargeted metabolomics**. *Analytical Chemistry* 2014, **86**(15): 7320-7327.

86. Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H, Harms AC, Ulrich EL, Markley JL, Sussman MR: **Stable Isotope Assisted Assignment of Elemental Compositions for Metabolomics**. *Analytical Chemistry* 2007, **79**(18): 6912-6921.

87. Nakabayashi R, Sawada Y, Yamada Y, Suzuki M, Hirai MY, Sakurai T, Saito K: **Combination of Liquid Chromatography Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry with C-13-Labeling for Chemical Assignment of Sulfur-Containing Metabolites in Onion Bulbs**. *Analytical Chemistry* 2013, **85**(3): 1310-1315.

88. De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD: **Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry**. *Nature Protocols* 2007, **2**(4): 778-791.

89. Hunter JD: **Matplotlib: A 2D graphics environment**. *Computing in Science & Engineering* 2007, **9**(3): 90-95.

90. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu WM, Aebersold R: **A common open representation of mass spectrometry data and its application to proteomics research**. *Nature Biotechnology* 2004, **22**(11): 1459-1466.

91. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Ropp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz PA, Deutsch EW: **mzML-a Community Standard for Mass Spectrometry Data**. *Molecular & Cellular Proteomics* 2011, **10**(1).

92. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, Eng JK, Martin DB, Nesvizhskii AI, Aebersold R: **A guided tour of the Trans-Proteomic Pipeline**. *Proteomics* 2010, **10**(6): 1150-1159.

93. Kessner D, Chambers M, Burke R, Agusand D, Mallick P: **ProteoWizard: open source software for rapid proteomics tools development**. *Bioinformatics* 2008, **24**(21): 2534-2536.

94. R Core Team: **R: A Language and Environment for Statistical Computing**. http://www.R-project.org. 2014.

95. Du P, Kibbe WA, Lin SM: **Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching**. *Bioinformatics* 2006, **22**(17): 2059-2065.

96. Sturm M, Kohlbacher O: **TOPPView: an open-source viewer for mass spectrometry data**. *Journal of Proteome Research* 2009, **8**(7): 3760-3763.

97. Lemmens M, Scholz U, Berthiller F, Dall'Asta C, Koutnik A, Schuhmacher R, Adam G, Buerstmayr H, Mesterhazy A, Krska R, Ruckenbauer P: **The ability to detoxify the mycotoxin deoxynivalenol colocalizes with a major quantitative trait locus for Fusarium head blight resistance in wheat**. *Molecular Plant-Microbe Interactions* 2005, **18**(12): 1318-24.

98. Ferrer-Gallego R, Hernandez-Hierro JM, Rivas-Gonzalo JC, Escribano-Bailon MT: **Influence of climatic conditions on the phenolic composition of *Vitis vinifera* L. cv. Graciano**. *Analytica Chimica Acta* 2012, **732**: 73-77.

99. Gunnaiah R, Kushalappa AC: **Metabolomics deciphers the host resistance mechanisms in wheat cultivar Sumai-3, against trichothecene producing and non-producing isolates of *Fusarium graminearum***. *Plant Physiol Biochem* 2014, **83C**: 40-50.

100. Brown M, Wedge DC, Goodacre R, Kell DB, Baker PN, Kenny LC, Mamas MA, Neyses L, Dunn WB: **Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets**. *Bioinformatics* 2011, **27**(8): 1108-1112.

101. Kaufmann A, Walker S: **Accuracy of relative isotopic abundance and mass measurements in a single-stage orbitrap mass spectrometer**. *Rapid Communications in Mass Spectrometry* 2012, **26**(9): 1081-1090.

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

103. Katajamaa M, Oresic M: **Data processing for mass spectrometry-based metabolomics**. *Journal of Chromatography A* 2007, **1158**(1-2): 318-328.

104. Antignac J, de Wasch K, Monteau F, De Brabander H, Andre F, Le Bizec B: **The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis**. *Analytica Chimica Acta* 2005, **529**(1–2): 129-136.

105. Warth B, Parich A, Bueschl C, Schoefbeck D, Nora N, Kluger B, Schuster K, Krska R, Adam G, Lemmens M, Schuhmacher R: **GC-MS based targeted metabolic profiling identified changes in the wheat metabolome following deoxynivalenol treatment**. *Metabolomics* 2014. doi: 10.1007/s11306-014-0731-1.

106. Bueschl C, Kluger B, Berthiller F, Lirk G, Winkler S, Krska R, Schuhmacher R: **MetExtract: a new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research**. *Bioinformatics* 2012, **28**(5): 736-738.

Part II

Original works

From the 8 publications, which were realised during the doctoral study of the author of this thesis, 5 were selected to form an integral part of the presented work.

The first publication (#1) was realised during an early stage of the author's doctoral study. This trend article summarises potential applications of SIL in untargeted metabolomics research, which include full metabolome labelling applications, tracer-fate studies, and derivatization with native and labelled agents. In that respect this publication and the preliminary version of MetExtract [106] served as a good basis for further work during the author's doctoral study.

Publication #3 and #5 detail the developed analytical workflows for full metabolome labelling and tracer-fate studies. Furthermore, they briefly summarise the steps for automated data processing of the acquired LC-HRMS data and present the accompanying metabolomics experiments and data processing results with a special focus laid on the benefits gained with SIL. The results show that SIL has many advantages over conventional, labelling-free approaches and also improves successive interpretation of single metabolites.

Publications #2 and #4 detail actual applications of the implemented workflows and developed software. The focus is put on the biological results of the performed experiments. In both publications previously unknown biotransformation products of the respectively studied tracer substance were detected. Furthermore, publication #5 includes biological results from a tracer-fate study.

The biological questions and workflows were developed in close cooperation between biologists, analytical chemists and bioinformaticians. In general, the author of this thesis was responsible for developing the software MetExtract and performing automatic data processing and statistical analysis of the generated metabolomics data. Moreover, the author wrote those parts of the manuscripts that detail the bioinformatics-related aspects and, together with the co-authors, wrote the remaining texts of the respective manuscripts.

Besides these contributions, the author of this thesis was also responsible for statistical investigation of GCMS metabolomics data presented by Warth and colleagues [105] and performed data processing of LC-HRMS data with MetExtract with the aim of finding suitable metabolites for demonstration of FragExtract [85]. Moreover, the author of this thesis also performed data processing and statistical investigation in currently not published metabolomics experiments.

In the following the 5 selected publications are reprinted and the contributions of the author of this thesis are detailed.

Publication #1: Isotopic labelling-assisted metabolomics using LC-MS

Christoph Bueschl, Rudolf Krska, Bernhard Kluger, Rainer Schuhmacher
Analytical and Bioanalytical Chemistry. 2013 Jan;405(1):27-33

Contributions of the presenting author: The author of this thesis formally defined the SIL experiments presented in this publication as well as their respective LC-HRMS data structures originating from the concurrent measurements of native and labelled metabolites. Additionally, SIL assisted data processing was performed to illustrate the characteristics of LC-HRMS data derived from SIL-assisted metabolomics experiments with a preliminary version of MetExtract. The author wrote the publication text and implemented the co-authors' feedback.

Anal Bioanal Chem (2013) 405:27–33
DOI 10.1007/s00216-012-6375-y

TRENDS

Isotopic labeling-assisted metabolomics using LC–MS

C. Bueschl · R. Krska · B. Kluger · R. Schuhmacher

Received: 6 July 2012 / Revised: 14 August 2012 / Accepted: 17 August 2012 / Published online: 26 September 2012

Abstract Metabolomics has emerged as the latest of the so-called "omics" disciplines and has great potential to provide deeper understanding of fundamental biochemical processes at the biological system level. Among recent technological developments, LC–HRMS enables determination of hundreds to thousands of metabolites over a wide range of concentrations and has developed into one of the most powerful techniques in non-targeted metabolomics. The analysis of mixtures of in-vivo-stable isotopic-labeled samples or reference substances with un-labeled samples leads to specific LC–MS data patterns which can be systematically exploited in practically all data-processing steps. This includes recognition of true metabolite-derived analytical features in highly complex LC–MS data and characterization of the global biochemical composition of biological samples. In addition, stable-isotopic labeling can be used for more accurate quantification (via internal standardization) and identification of compounds in different organisms.

Keywords Bioanalytical methods · Mass spectrometry · Metabolomics · Liquid chromatography

Metabolomics: a brief introduction

The objective of metabolomics is comprehensive, qualitative and quantitative analysis of all the low-molecular-weight metabolites of a living cell, organ, or whole organism [1]. The term "metabolome" has been defined by analogy with the genome and refers to the complete set of metabolites of a biological system [2, 3]. Thus metabolomics can be regarded as characterization of the metabolome. Although genomes have been sequenced for many organisms, it is currently not possible to measure the whole metabolome of a biological system at once, because of analytical–methodical limitations and the highly diverse nature of the metabolites.

In this respect, two different metabolomics concepts can be distinguished: targeted and non-targeted. In targeted approaches, abundances of metabolites of a set of predefined known substances are determined. Such an approach enables absolute quantification but is usually limited to metabolites which are available as authentic reference standards. In contrast, non-targeted approaches try to find mass spectrometric features of all detectable compounds, including those unknown at the time of sample measurement. This approach therefore has the advantage of probing the entire metabolic space and can obtain relative abundances of several hundreds to thousands of known and unknown metabolites [4].

Currently, most non-targeted metabolomics studies use liquid chromatography coupled to high-resolution mass spectrometry (LC–HRMS). This combination enables the detection of the highest number of metabolites and requires only small amounts of the biological sample [5]. The combination of electrospray ionization (ESI) with full-scan LC–MS at unit resolution results in a large number of solvent-related cluster ions and other non-metabolite-related signals which might interfere with masses of true metabolite ions. For this reason, low resolution ESI LC–MS is usually restricted to targeted approaches in which MS–MS modes are mostly applied. Most of the applications described in this article were, however, used for non-targeted analysis of biological samples using full-scan LC–HRMS.

Sample measurements with modern analytical instrumentation result in huge amounts of data, which can no longer be evaluated manually. Data-processing steps serve to

C. Bueschl · R. Krska · B. Kluger · R. Schuhmacher (✉)
Center for Analytical Chemistry,
Department for Agrobiotechnology (IFA-Tulln),
University of Natural Resources and Life Sciences, Vienna,
Konrad-Lorenz-Str 20,
3430 Tulln, Austria
e-mail: rainer.schuhmacher@boku.ac.at

reduce data complexity and include numerous elements, for example feature extraction, spectrum deconvolution (i.e., grouping of ions which originate from the same metabolite), retention time (R_t) alignment of chromatographic peaks across different runs, and internal standardization for quantification. The final result of data processing is a data matrix, a table that contains all samples and analytical features (i.e., metabolite-derived signals). The ultimate objective of a metabolomics study is to detect as many metabolites as possible and to link the differently expressed metabolites to a variety of experimental factors. Therefore, reliable annotation and/or identification of the detected metabolites is essential.

Despite the availability of several software tools for data processing [6–11] and improved measurement techniques, several limitations remain.

From the extraction of metabolite-derived analytical features to the annotation and identification of metabolites

A first step of data processing in non-targeted metabolomics approaches is to extract as many features as possible. The term "feature" denotes a two dimensional bounded signal consisting of a chromatographic peak (i.e., retention time) and an MS peak (i.e., *m*/*z* value) [12]. The comprehensive and reliable extraction of metabolite-derived features of true biological origin however, remains a very difficult task. A major limitation is the large amount of non-metabolite-related noise and background signals. It has been estimated that in LC–electrospray ionization–MS (LC–ESI–MS), as little as 10 % of the signals are of true biological origin [13]. Consequently, most features are not associated with true metabolites but can hardly be recognized as such, because their analytical characteristics are the same as for features of true biological origin. Furthermore, a single metabolite leads to more than one ion species—e.g., isotopic peaks, adducts, in-source fragments. Therefore, data processing requires the assignment of metabolite ions and the grouping of features into deconvoluted spectra before meaningful annotation of metabolites can be achieved. Typically, feature abundances are estimated by peak integration and the resulting peak areas are used for relative quantification and comparison of different experimental states.

After feature extraction and spectral deconvolution, annotation and/or identification of metabolites is essential, but is one of the most challenging tasks of any metabolomics experiment [14–16]. Metabolite annotation usually uses a search for one or more molecular properties (e.g., accurate mass, sum formula) against comprehensive databases (DB), whereas identification also requires confirmation by measurement of an authentic standard under identical analytical conditions and/or comparison of MS–MS fragmentation patterns.

DB search approaches seldom provide unique identifications, since usually more than one substance exists for a particular mass or sum formula, because of the great number of combinations of one sum formula into numerous structural formulas. Furthermore, because non-targeted metabolomics experiments yield both known and unknown metabolites, it can be expected that a significant number of metabolites is missing from the DBs and thus remains uncharacterized [5]. Additional MS–MS measurements are required to provide structure information. Unfortunately, MS–MS measurements are very labor-intensive tasks even for only a small subset of putative metabolites and the resulting MS–MS spectra are difficult to interpret. Furthermore, for LC–MS very few databases with authentic MS–MS spectra are available, again reducing the number of identifiable metabolites.

With regard to quantification of metabolites by LC–ESI–MS, matrix effects, or signal suppression or enhancement (SSE) (for recent review see, e.g., Refs. [17] and [18]), limit the accuracy and reliability of quantitative measurements within and between different measurement sequences. SSE is caused by the presence of (endogenous or exogenous) co-eluting components in the ion source of the mass spectrometer and has been attributed to numerous mechanisms including competition for "charges" between analytes and interfering compounds or a change of viscosity and/or surface tension of the droplets in the ion source [17, 18].

The potential of stable isotopic labeling to meet current challenges in LC–MS-based metabolomics research

In view of these limitations, there is a strong need both for innovative approaches to the analytical measurement of biological samples and for the development of novel, improved data-processing algorithms and their implementation in the form of user-friendly software tools, especially for non-targeted metabolomics.

In this respect, stable isotopic labeling (SIL) is a very promising and increasingly popular technique in metabolomics research, which is perfectly suited to be combined with GC–MS and LC–MS. Stable isotope techniques were preceded by the use of radioactive isotopes in biochemical research. The development of robust and sensitive GC–MS and LC–MS instrumentation together with easier and safer handling procedures (i.e., organizational restrictions, human health concerns), quickly increased the use of radio isotopes [19]. First applications of SIL in proteomics and protein labeling demonstrated the huge advantages of this technique [20] and researchers therefore adopted SIL in a variety of metabolomics techniques. In this respect ^{13}C is the most commonly used isotope for SIL-assisted experiments, because ^{13}C-labeled isotopologues cannot be chromatographically separated from their natural analogues (Fig. 1a).

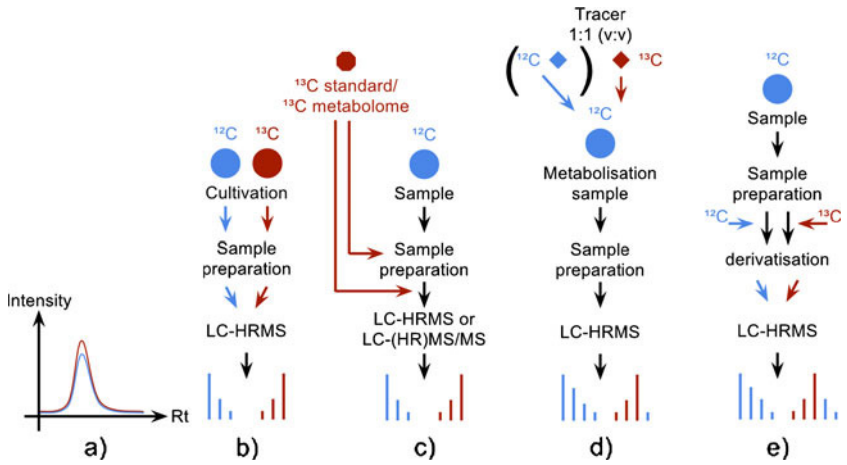


Fig. 1 Different strategies for using SIL to assist metabolomics studies. **a** ^{13}C, ^{15}N, and ^{34}S-enriched substances are not chromatographically separated from the corresponding natural isotopologues, thus the non-labeled and the labeled isotopologues elute at the same retention time with identical peak profiles. **b** For non-targeted annotation of an organism's metabolome the organism can be cultivated in parallel using differently isotopologue-enriched nutrition sources (e.g., ^{12}C and ^{13}C glucose as sole carbon source). The extracts are subsequently mixed and measured with LC–HRMS. The resulting data pattern helps in the extraction of true biological signals. **c** Absolute compound quantification using an authentic, labeled standard or relative quantification using a stock of globally labeled sample extract of the same organism for inter-experiment comparison. **d** Metabolism experiment using natural and fully labeled tracer substances enables metabolism studies and greatly helps to separate products of metabolism from other biological signals. In contrast with metabolism studies, fluxomics experiments only spike with the labeled tracer. **e** Derivatisation using non-labeled and labeled derivatisation agents enables rapid recovery of many metabolites belonging to the same chemical groups (e.g., alcohols, acids …)

Moreover, carbon is a constituent of every metabolite and its transfer between biological entities follows well known rules [19]. Although ^{15}N and ^{34}S are also well suited to investigation of all nitrogen or sulfur-containing compounds and their bio-transformation by living systems, ^{2}H and ^{18}O are less frequently used for labeling of metabolites in metabolomics, because they frequently tend to be exchanged by hydrogen and oxygen from surrounding water. Moreover deuterated metabolites do not always perfectly co-elute with their non-labeled analogues [19], rendering the extent of labeling and the results obtained from subsequent data processing difficult to interpret. Figure 2 shows a three dimensional view of an LC–HRMS chromatogram of the supernatant from the agricultural important fungus *F. graminearum*, which was cultivated in parallel on ^{12}C and ^{13}C glucose medium. Identical aliquots of the ^{12}C and ^{13}C supernatant were mixed and subsequently measured using an LTQ Orbitrap XL. As one can see in the circled area, the LC elution pattern for the ^{12}C (natural) and ^{13}C-labeled metabolite ions of the same metabolite are identical but have a defined *m/z* shift (861.3835 for the ^{12}C ion and 897.5043 for the corresponding ^{13}C ion). Because ^{12}C and ^{13}C-labeled glucose were used for cultivation, the number of carbon atoms in this particular ion can be calculated from the *m/z* difference between the two monoisotopic mass peaks. For this example, the *m/z* difference of 36.1208 suggests 36 carbon atoms for this putative metabolite.

To further confirm that the two isotopologue ions belong to the same putative metabolite, their carbon-isotopic distributions are of interest. Whereas descending isotopic distributions towards higher *m/z* values (the abundance of natural occurring ^{12}C is approximately 98.93 % and thus the abundance of ^{13}C is 1.07 %) are observed for all natural ^{12}C substances, ^{13}C-labeled substances do not have descending isotopic distributions, because their most abundant carbon isotope is ^{13}C and only the minority of carbon atoms are ^{12}C isotopes (the abundance of ^{13}C and ^{12}C is highly dependent on the enrichment method and cannot be generalized in the same way as natural carbon abundances). Thus ^{13}C-labeled monoisotopic ions are not only heavier but also have ascending isotopic distribution toward higher *m/z* values, because ^{13}C isotopes are replaced by the lighter ^{12}C isotopes. These two mirror symmetric isotopic distributions, which are only present for SIL using carbon, confirm that both ions are successors of the same metabolite and enable the number of carbon atoms to be easily determined. Depending on the MS instrument used, the relative abundance of the isotopic distributions are more or less accurately recorded. In general, TOF instruments yield more accurate relative abundances of isotopic peaks than Orbitrap or other FT mass analyzers, whereas the resolving power and thus mass accuracy are higher with FT-MS instruments. The theoretical isotopic-abundance pattern for a specific isotopologue in comparison with the substance's most intense isotopic peak is calculated by use of the formula:

$$P\left(\vec{a},\vec{s},\vec{p},\vec{p_e}\right)=\prod_{i=0}^{|a|}\frac{{p_{ei}}^{a_i-s_i}{p_i}^{S_i}\binom{a_i}{s_i}}{{p_{ei}}^{a_i}}$$

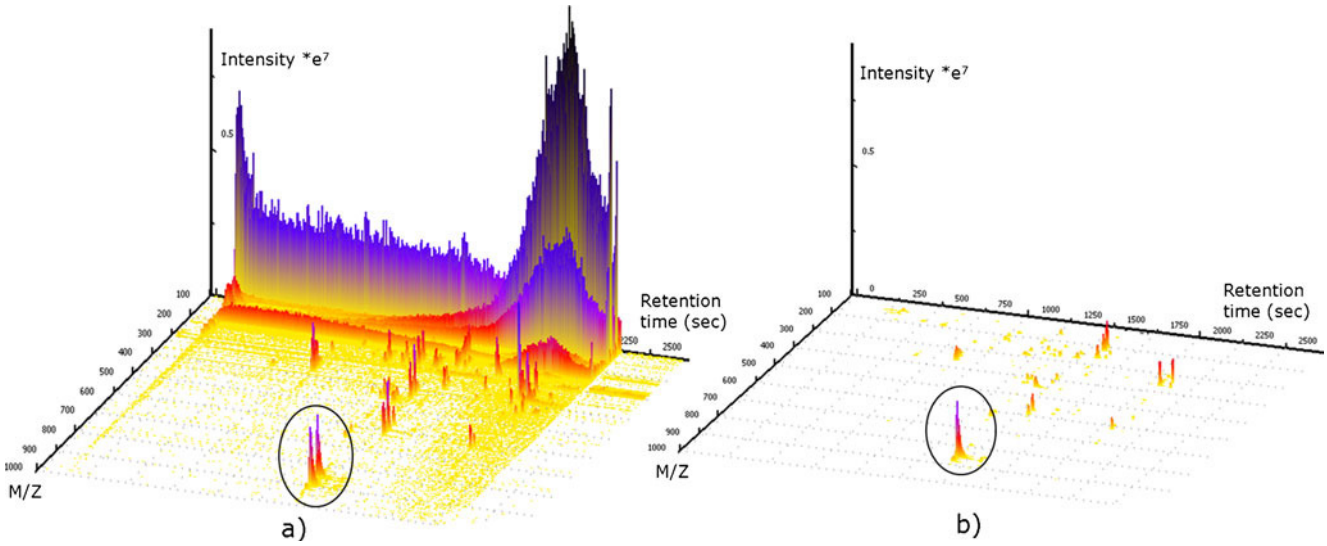


Fig. 2 3D Plots obtained from *F. graminearum*. **a** Unprocessed LC–HRMS full-scan chromatograms of a mixture of supernatants from cultivation of *F. graminearum* on both ^{12}C and ^{13}C glucose. The *circle* marks an ion pair (the ^{12}C and the corresponding ^{13}C-labeled monoisotopic ions) originating from the same metabolite. After processing the spectrum with MetExtract, only ions having a labeled pendant are not removed (**b**). Thus, only the non-labeled ions remain in the circled area. (3D view generated with Ref. [37])

where a_i is the number of atoms of the *i*th element in the substance, s_i is the number of substitutions of the most abundant isotopologue with less abundant isotopes of the *i*th element, p_i is the isotopic purity of the used stable isotope of the *i*th element, and p_{ei} is the isotopic purity of the most abundant isotope of the *i*th element (e.g., the relative abundance of the first isotopologue ($^{13}C_1$) compared with its ^{12}C monoisotopic peak for an ion having 24 carbon atoms is P([24], [1], [0.0107], [0.9893])=25.68 %. The same isotopologue having additionally replaced one of nine oxygen atoms by ^{17}O (natural abundances for ^{16}O and ^{17}O are 99.962 % and 0.038 %) is P([9, 24], [1, 1], [0.0107, 0.00038], [0.9893, 0.99962])= 0.08 %). However, very low isotopic abundances (e.g., ^{17}O, ^{36}S, ^{2}H …) may be of theoretical interest but, in general, do not have to be considered practically in LC–HRMS analysis because they are not observable with most current MS instrumentation.

Furthermore, the mirror symmetric isotopic distributions in the case of carbon SIL are very helpful if the substances of interest contain heteroatoms such as sulfur, nitrogen, or calcium. Without the SIL process, heteroatom isotopologues cannot usually be separated from the more intense ^{13}C isotopologues of the natural, ^{12}C ions even at a high MS resolving power of 100,000. With carbon SIL additional isotopic mass peaks originating from heteroatom isotopes (e.g., ^{15}N, ^{34}S) can be observed next to the ^{13}C ion at higher *m/z* values. This benefit is very interesting for non-targeted metabolomics because it provides further information about the elemental composition of the detected ion species.

Apart from non-targeted metabolomics, feature extraction and annotation, SIL is also perfectly suited to absolute quantification. Because the natural and spiked labeled reference standard perfectly co-elute, the ratio of their peak areas is used for absolute quantification. Furthermore, because the standard is added to the biological sample, its concentration is known. Using the ratio of the two isotopologue peak areas and the amount of reference standard, absolute quantification of the natural isotopologue in the sample can be achieved.

Examples of applications of SIL

All of the following SIL-assisted approaches involve analysis of mixtures of labeled and non-labeled samples and are based on enrichment of the respective mass spectra with the isotopic pattern of the SIL analogue(s). Depending on the purpose of the labeling experiment (Fig. 1b–e), the degree of isotopic enrichment, metabolic fate of the tracer(s), and other factors, the resulting mass spectra might be affected in different ways but all have the SIL-specific data pattern exemplified in Fig. 1. The mass spectra might just contain (slightly) altered relative intensities of individual isotopologues or pairs of clearly separated isotopic patterns of the natural compounds and their corresponding fully labeled analogues.

In the following four main areas involving SIL and LC–MS in metabolomics research will be presented with the focus on LC–HRMS.

SIL-assisted whole metabolome studies (Fig. 1b)

For reliable extraction and annotation of putative metabolites in a biological sample, non-labeled and labeled cultures of the organism of interest are mixed and measured by use

of LC–HRMS in the full-scan mode. Whole labeled metabolomes of specific organisms can be obtained when a fully labeled substrate, for example $^{13}C_6$ glucose forms the sole carbon source of, e.g., bacteria or filamentous fungi. Such in vivo labeling of microbes has been successfully used to accomplish internal standardization for quantification, to circumvent problems of signal suppression in ESI–MS [21]. For this purpose, in-vivo-labeled samples can either be cultured in parallel to every experimental sample and condition and subsequently mixed with non-labeled analogues, or prepared by producing a large quantity of labeled culture with which non-labeled experimental samples can be spiked [22, 23].

Recently, in-vivo labeling has also been extended to plants and used to facilitate the assignment of the elemental composition of metabolite ions by accurate mass measurements in combination with database search [24, 25]. Moreover, as we have demonstrated in our own recent work, in-vivo labeling of fungal culture samples provides a powerful approach for the automated global extraction of all metabolite-derived MS signals from LC–HRMS raw data by discrimination of true metabolite-related from non-specific analytical features [26].

SIL improves absolute quantification (Fig. 1c)

In targeted analysis, SIL reference standards can be used for internal standardization of peak intensities in trace analysis by use of LC–HRMS or LC–(HR)MS–MS [27, 28], making this the most common use of SIL. Use of authentic, labeled reference standards enables easy, rapid, and absolute quantification of the respective metabolites even in complex matrixes.

In vivo metabolism studies (Fig. 1d)

In non-targeted metabolomics, uniformly labeled substances in combination with their natural pendants are added to biological samples during cultivation and used as tracers to study their metabolic fate [4]. This technique has been used in targeted approaches to study metabolic pathways and fluxes of the central metabolism [29, 30]. Moreover, labeled tracers have also been used to study the bio-transformation of secondary metabolites [31] and xenobiotics [32]. Without use of SIL for this kind of research, searching for and detecting these products of metabolism in complex full-scan LC–HRMS data by non-targeted approaches would be almost impossible. Without the use of SIL one would have to subtract chromatograms without the spiked compound from those obtained after measurement of samples with the specific compounds added. Although possible in principle, the latter approach results in numerous false-positive findings because of fluctuations in MS signals from measurement to measurement.

SIL-assisted derivatisation (Fig. 1e)

Another recently introduced SIL-based technique in combination with LC–HRMS is the use of labeled reagents for derivatisation which enables non-targeted screening for all compounds belonging to a specific chemical group (e.g., alcohols, acids …). The general workflow for a SIL derivatisation step is to split the biological sample into two identical aliquots and perform derivatisation separately with labeled and non-labeled derivatisation reagents, and mix and measure them jointly. Thus, one not only gains all the benefits derivatisation has but also the benefits of SIL which, in this case, are improved metabolite feature extraction in the highly complex LC–HRMS data [33, 34] and an estimate of the exchanged functional groups.

Despite the high potential of labeling-assisted metabolomics approaches, only a few data processing tools have been published which specifically exploit the labeling-associated data pattern. Moreover, these tools are limited to targeted metabolomics and fluxomics approaches. For targeted GC–MS-based metabolomics of cell cultures, Hiller and colleagues published a method for study of the fate of labeled tracer compounds through central metabolism [35]. For targeted LC–MS studies, commercial software (IROA) exists which has been designed to use mixtures of in-vivo labeled and non-labeled biological culture samples to assign differently expressed metabolites in differently treated biological samples. The software calculates and graphically illustrates the intensity ratios of the principal ions from predefined isotopologue signal pairs [36]. Another software product also utilizing ^{13}C labeling is MetMax [11], which has been used to analyze the dynamics of CO_2 uptake by *Chlamydomonas reinhardtii* using GC × GC–TOF-MS. To the best of our knowledge, the recently developed MetExtract software is, to date, the only publicly available tool for non-targeted, automated global detection of metabolite-derived LC–HRMS signals originating from natural and stable isotopically labeled analogues and their assignment to true biological metabolites [26].

Although LC–MS experiments using stable isotopically labeled compounds in combination with their naturally occurring pendants is already a well established and frequently used technique, the most limiting disadvantage is the relatively high cost associated with enrichment of compounds with some stable isotopes. Furthermore, a stable isotopic labeled source which can be used for in-vivo labeling experiments or as an internal standard may not be available commercially. Authentic labeled reference standards or nutrition sources have to be synthesized or harvested from organisms and subsequently purified, which again is a cost-intensive exercise. For reference standards, the second major limitation is availability because only a very small subset of all needed reference standards are available, which limits their advantage. Another limitation of

SIL and whole-metabolome experiments is the requirement to cultivate the organisms in parallel—once with the natural and once with the stable isotope enriched nutrition source—to be able to recover experiment or condition-specific metabolites of the organism of interest.

Outlook

Although SIL-based analytical approaches have been developed and frequently used in both proteomics and biochemical research for many years, the metabolomics community started to fully exploit the potential of this technique only recently. Considering the numerous benefits and advantages of SIL, we expect its increased application, particularly for non-targeted metabolomics research. Although stable isotopically labeled compounds or biological samples are quite expensive, they offer many benefits for analytical chemists working with LC–MS. The unique and undistinguishable data pattern obtained from natural and fully labeled isotopologues of a compound drastically simplifies data processing. Furthermore, in target analysis the addition of labeled authentic standards enables absolute quantification of substances in biological samples. For non-targeted metabolomics SIL-based LC–HRMS approaches enable global data extraction and feature annotation for true metabolites and can help to improve precision in relative quantification. With further improved analytical instrumentation and customized data-processing software for SIL-derived data patterns, this technique has the potential to be of crucial importance in the new discipline of metabolomics.

Acknowledgments The authors thank the Austrian Science Fund (project SFB Fusarium 3706-B11) and the WWTF (project ToxiGenome 9793008037) for financial support.

References

1. Fiehn O (2002) Metabolomics–the link between genotypes and phenotypes. Plant Mol Biol 48(1–2):155–171
2. Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16 (9):373–378
3. Weckwerth W (2010) Metabolomics: an integral technique in systems biology. Bioanalysis 2(4):829–836. doi:10.4155/bio.09.192
4. Hiller K, Metallo C, Stephanopoulos G (2011) Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics. Curr Pharm Biotechnol 12(7):1075–1086
5. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13(4):263–269. doi:10.1038/nrm3314
6. Tautenhahn R, Boettcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. BMC Bioinforma 9:504. doi:10.1186/1471-2105-9-504
7. Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinforma 11:395. doi:10.1186/1471-2105-11-395
8. Kohlbacher O, Reinert K, Groepl C, Lange E, Pfeifer N, Schulz-Trieglaff O et al (2007) TOPP - the OpenMS proteomics pipeline. Bioinformatics 23(2):e191–e197. doi:10.1093/bioinformatics/btl299
9. Lommen A (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem 81(8):3079–3086. doi:10.1021/ac900036d
10. Melamud E, Vastag L, Rabinowitz JD (2010) Metabolomic analysis and visualization engine for LC–MS data. Anal Chem 82 (23):9818–9826. doi:10.1021/ac1021166
11. Kempa S, Hummel J, Schwemmer T, Pietzke M, Strehmel N, Wienkoop S et al (2009) An automated GCxGC–TOF-MS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential 13C-labelling experiments: a case study for photoautotrophic-mixotrophic grown Chlamydomonas reinhardtii cells. J Basic Microbiol 49(1):82–91. doi:10.1002/jobm.200800337
12. Kuhl C, Tautenhahn R, Boettcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84(1):283–289. doi:10.1021/ac202450g
13. Keller BO, Sui J, Young AB, Whittal RM (2008) Interferences and contaminants encountered in modern mass spectrometry. Anal Chim Acta 627(1):71–81. doi:10.1016/j.aca.2008.04.043
14. Hall RD (2011) Plant metabolomics in a nutshell: potential and future challenges Wiley–Blackwell, pp 1–24)
15. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B et al (2009) Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 5(4):435–458. doi:10.1007/s11306-009-0168-0
16. Zhou B, Xiao JF, Tuli L, Ressom HW (2012) LC–MS-based metabolomics. Mol Biosyst 8(2):470–481. doi:10.1039/c1mb05350g
17. Antignac J-P, de Wasch K, Monteau F, Brabander HD, Andre FO, Bizec BL (2005) The ion suppression phenomenon in liquid chromatography mass spectrometry and its consequences in the field of residue analysis. Anal Chim Acta 529(12):129–136. doi:10.1016/j.aca.2004.08.055
18. Niessen WMA, Manini P, Andreoli R (2006) Matrix effects in quantitative pesticide analysis using liquid chromatography–mass spectrometry. Mass Spectrom Rev 25(6):881–899. doi:10.1002/mas.20097
19. Klein S, Heinzle E (2012) Isotope labeling experiments in metabolomics and fluxomics. Wiley Interdiscip Rev Syst Biol Med 4 (3):261–272. doi:10.1002/wsbm.1167
20. Beynon RJ, Pratt JM (2005) Metabolic labeling of proteins for proteomics. Mol Cell Proteomics 4(7):857–872. doi:10.1074/mcp.R400010-MCP200
21. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc 3(8):1299–1311. doi:10.1038/nprot.2008.107
22. Hegeman AD (2010) Plant metabolomics–meeting the analytical challenges of comprehensive metabolite analysis. Brief Funct Genomics 9(2):139–148. doi:10.1093/bfgp/elp053
23. Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C et al (2005) Quantitative analysis of the microbial metabolome by

isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards. Anal Biochem 336 (2):164–171. doi:10.1016/j.ab.2004.09.001
24. Giavalisco P, Koehl K, Hummel J, Seiwert B, Willmitzer L (2009) ^{13}C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography–mass spectrometry-based metabolomic research. Anal Chem 81(15):6546–6551. doi:10.1021/ac900979e
25. Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H et al (2007) Stable isotope assisted assignment of elemental compositions for metabolomics. Anal Chem 79(18):6912–6921. doi:10.1021/ac070346t
26. Bueschl C, Kluger B, Berthiller F, Lirk G, Winkler S, Krska R et al (2012) MetExtract: a new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. Bioinformatics 28(5):736–738. doi:10.1093/bioinformatics/bts012
27. Sulyok M, Berthiller F, Krska R, Schuhmacher R (2006) Development and validation of a liquid chromatography/tandem mass spectrometric method for the determination of 39 mycotoxins in wheat and maize. Rapid Commun Mass Spectrom 20(18):2649–2659. doi:10.1002/rcm.2640
28. Berthiller F, Sulyok M, Krska R, Schuhmacher R (2007) Chromatographic methods for the simultaneous determination of mycotoxins and their conjugates in cereals. Int J Food Microbiol 119(1–2):33–37. doi:10.1016/j.ijfoodmicro.2007.07.022
29. Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J (2005) Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends Biotechnol 23(1):28–33. doi:10.1016/j.tibtech.2004.12.001
30. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ et al (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nat Genet 43 (9):869–874. doi:10.1038/ng.890
31. Chen C, Gonzalez FJ, Idle JR (2007) LC–MS-based metabolomics in drug metabolism. Drug Metab Rev 39(2–3):581–597. doi:10.1080/03602530701497804
32. Athersuch TJ, Nicholson JK, Wilson ID (2007) Isotopic enrichment enhancement in metabonomic analysis of UPLC–MS data sets. J Labelled Comp Radiopharm 50(5–6):303–307. doi:10.1002/Jlcr.1217
33. Guo K, Li L (2009) Differential ^{12}C-/^{13}C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. Anal Chem 81(10):3919–3932. doi:10.1021/ac900166a
34. Guo K, Bamforth F, Li L (2011) Qualitative metabolome analysis of human cerebrospinal fluid by 13C-/12C-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom 22(2):339–347. doi:10.1007/s13361-010-0033-4
35. Hiller K, Metallo CM, Kelleher JK, Stephanopoulos G (2010) Nontargeted elucidation of metabolic pathways using stable-isotope tracers and mass spectrometry. Anal Chem 82(15):6621–6628. doi:10.1021/ac1011574
36. Beecher (2010) Method for generation and use of isotopic patterns in mass spectral data of simple organisms. U.S. patent 7,820,963. In I. Metabolic Alayses (Ed)
37. Sturm M, Kohlbacher O (2009) TOPPView: an open-source viewer for mass spectrometry data. J Proteome Res 8(7):3760–3763. doi:10.1021/pr900171m

Publication #2: Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat

Bernhard Kluger, **Christoph Bueschl**, Marc Lemmens, Franz Berthiller, Georg Häubl, Günther Jaunecker, Gerhard Adam, Rudolf Krska, Rainer Schuhmacher
Analytical and Bioanalytical Chemistry. 2013 Jun;405(15):5031-5036

Contributions of the presenting author: The author of this thesis extended MetExtract to support the detection of biotransformation products derived from native and labelled tracer substances. Additionally, the author of this thesis performed automatic data processing with MetExtract of the wheat samples treated with native and globally ^{13}C-labelled DON and assisted in preparing the illustrations in the manuscript.

Anal Bioanal Chem (2013) 405:5031–5036
DOI 10.1007/s00216-012-6483-8

SHORT COMMUNICATION

Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat

Bernhard Kluger · Christoph Bueschl · Marc Lemmens · Franz Berthiller · Georg Häubl · Günther Jaunecker · Gerhard Adam · Rudolf Krska · Rainer Schuhmacher

Received: 20 August 2012 / Revised: 28 September 2012 / Accepted: 5 October 2012 / Published online: 20 October 2012

Abstract An untargeted screening strategy for the detection of biotransformation products of xenobiotics using stable isotopic labelling (SIL) and liquid chromatography–high resolution mass spectrometry (LC-HRMS) is reported. The organism of interest is treated with a mixture of labelled and non-labelled precursor and samples are analysed by LC-HRMS. Raw data are processed with the recently developed MetExtract software for the automated extraction of corresponding peak pairs. The SIL-assisted approach is exemplified by the metabolisation of the *Fusarium* mycotoxin deoxynivalenol (DON) *in planta*. Flowering ears were inoculated with 100 μg of a 1+1 (*v*/*v*) mixture of non-labelled and fully labelled DON. Subsequent sample preparation, LC-HRMS measurements and data processing revealed a total of 57 corresponding peak pairs, which originated from ten metabolites. Besides the known DON and DON-3-glucoside, which were confirmed by measurement of authentic standards, eight further DON-biotransformation products were found by the untargeted screening approach. Based on a mass deviation of less than ±5 ppm and MS/MS measurements, one of these products was annotated as DON-glutathione (GSH) conjugate, which is described here for the first time for wheat. Our data further suggest that two DON-GSH-related metabolites, the processing products DON-*S*-cysteine and DON-*S*-cysteinyl-glycine and five unknown DON conjugates were formed *in planta*. Future MS/MS measurements shall reveal the molecular structures of the detected conjugates in more detail.

Keywords Metabolisation · Xenobiotics · Stable isotopic labelling · Deoxynivalenol · Liquid chromatography–high resolution mass spectrometry · Mycotoxin conjugate

Published in the topical collection *Metabolomics and Metabolite Profiling* with guest editors Rainer Schuhmacher, Rudolf Krska, Roy Goodacre and Wolfram Weckwerth.

B. Kluger · C. Bueschl · F. Berthiller · R. Krska · R. Schuhmacher (✉)
Center for Analytical Chemistry, Department for Agrobiotechnology (IFA-Tulln), University of Natural Resources and Life Sciences (BOKU), Vienna,
Konrad-Lorenz-Str. 20,
3430 Tulln, Austria
e-mail: rainer.schuhmacher@boku.ac.at

M. Lemmens
Institute for Biotechnology in Plant Production, Department for Agrobiotechnology (IFA-Tulln), University of Natural Resources and Life Sciences (BOKU), Vienna,
Konrad-Lorenz-Str. 20,
3430 Tulln, Austria

G. Häubl · G. Jaunecker
Romer Labs Diagnostic GmbH,
Technopark 1,
3430 Tulln, Austria

G. Adam
Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences (BOKU), Vienna,
Konrad-Lorenz-Str. 24,
3430 Tulln, Austria

Introduction

Xenobiotics are frequently metabolised and subsequently conjugated to more polar derivatives as a part of detoxification strategies of organisms, including animals and plants. For the determination of metabolic pathways, mass spectrometry turned out to be one of the most powerful techniques for the detection of all types of low-molecular weight metabolites even on systems level [1–3]. In metabolomics, targeted approaches aim at the quantification of known or predicted specific metabolites, while untargeted approaches try to probe the global metabolic space of a biological sample [4]. As a major advantage over targeted methods,

untargeted metabolomics approaches have the potential to discover novel and unexpected metabolites such as unknown biotransformation products originating from specific xenobiotics.

Stable isotopic labelling (SIL)-assisted metabolomics techniques [5–9] offer a wide range of new applications in the field of untargeted approaches. In vivo SIL for example can be used for the global detection of biologically derived metabolite signals and offers a high degree of certainty that detected MS signals represent actual metabolites and are not resulting from e.g. solvents, matrix background or noise. Moreover, SIL-assisted approaches provide additional information on the number of carbon atoms of the detected metabolite ions [5]. Both stable isotopic labelled endogenous [10] as well as xenobiotic [3, 11] precursors have been used to study in vivo metabolisation in biological samples. In this context, a major challenge is the detection of the labelled biotransformation products within highly complex liquid chromatography–high resolution mass spectrometry (LC-HRMS) chromatograms. The recently developed software algorithm MetExtract [12] is capable of the automated global detection of metabolite-derived LC-HRMS signals originating from non-labelled and stable isotopically labelled compounds.

Here, we present the first results of a SIL-assisted metabolomics approach for the untargeted screening of xenobiotics and its derived biotransformation products exemplified by the metabolisation of the *Fusarium graminearum* mycotoxin deoxynivalenol (DON). DON [13] belongs to the group of trichothecene mycotoxins and is a frequent contaminant of food and feed. It is produced by *F. graminearum* during infection of cereal grains such as wheat and barley [14]. The protein synthesis inhibitor DON is a virulence factor of *Fusarium*, and in turn, detoxification of DON seems to be an important component of *Fusarium* resistance in wheat [15, 16]. Plants can reduce the toxicity of DON by conjugation of the toxin to glucose [17]. In this study, we applied a SIL-assisted approach for the untargeted screening of biotransformation products of DON in wheat. We found eight novel wheat-derived DON conjugates and for the first time obtained clear evidence that glutathione-mediated metabolism of DON occurs in wheat.

Experimental section

Chemicals and reagents

Methanol (MeOH, LiChrosolv, LC gradient grade) was purchased from Merck (Darmstadt, Germany); acetonitrile (ACN, HiPerSolv Chromanorm, HPLC gradient grade) was purchased from VWR (Vienna, Austria); formic acid (FA, MS grade) was obtained from Sigma-Aldrich (Vienna, Austria). Water was purified successively by reverse osmosis and an ELGA Purelab Ultra-AN-MK2 system (Veolia Water, Vienna, Austria). Non-labelled DON and uniformly labelled $^{13}C_{15}$ DON were obtained as a contribution for that particular study from Romer Labs (Tulln, Austria) in crystalline form, not commercially available. DON-3-glucoside (D3G) standard was produced according to [18].

Standard preparation

Standard solutions of DON and D3G with concentrations of 1 $mg L^{-1}$ were prepared in methanol/water (1+1, *v*/*v*) and used as analytical standards. Lyophilised non-labelled DON and $^{13}C_{15}$-labelled DON for inoculation of wheat were dissolved separately in pure water to a concentration of 1 $g L^{-1}$. Both stock solutions were mixed 1+1 (*v*/*v*) to obtain an inoculation solution containing 500 $mg L^{-1}$ non-labelled DON + 500 $mg L^{-1}$ $^{13}C_{15}$ DON.

Treatment of wheat samples

Wheat plants (cultivar “Remus”, which is sensitive for both Fusarium head blight and DON) were grown under standardised conditions [15]. Flowering ears were either treated with the DON inoculation solution or with water (mock) according to the following procedure: at time point zero, 10 μL DON were injected in each of two adjacent spikelets in the lower part of a flowering ear. Twenty-four hours later, the next two adjacent spikelets, located immediately above those previously treated, were injected with the same amount of toxin. This procedure was repeated 48, 72 and 96 h after the first treatment resulting in a total applied amount of 100 μg DON per ear. All treatments were carried out in triplicate. One hundred and eight hours after the first inoculation, the ear bases, containing all inoculated spikelets of DON-treated and mock ears, were sampled separately and immediately shock-frozen in liquid nitrogen. That way, we aimed to target all DON metabolites present *in planta* between 12 to 108 h post-treatment. All samples were stored at −80 °C until further sample preparation.

Sample preparation

Frozen wheat ears were milled separately to a fine powder for 2 min at 30 Hz using a ball mill (MM301 Retsch, Haan, Germany) with pre-cooled (liquid nitrogen) 10-mL stainless steel vessels (Retsch) and a 9-mm stainless steel ball (Retsch). Of homogenised wheat material, 105±5 mg was weighed to 1.5-mL Eppendorf tubes and extracted with 1 mL of pre-cooled methanol/water 3+1 *v*/*v* including 0.1 % formic acid, by vortexing for 10 s, and subsequent treatment in an ultrasonic bath at room temperature for 15 min according to de Vos et al. [19]. Samples were centrifuged for

4 min at 8,500×*g* (9,500 rpm) at room temperature. An aliquot of 600 μL supernatant was transferred to another 1.5-ml Eppendorf tube adding 300 μL water + 0.1 % formic acid to achieve a final methanol/water ratio of 1:1 (*v*/*v*). All samples were vortexed for 10 s before transfer into HPLC vials for LC-MS measurements.

LC-HR-MS conditions

LC-HR-MS measurements were performed on an LTQ Orbitrap XL (Thermo Fisher Scientific) equipped with an electrospray ionisation (ESI) source coupled to an UHPLC system (Accela, Thermo Fisher Scientific, San Jose, CA, USA). The analytical column was a reversed-phase XBridge C_{18}, 150 × 2.1 mm i.d., 3.5 μm particle size (Waters, Milford, MA, USA), preceded by a C_{18} 4 × 3 mm i.d. security cartridge (Phenomenex, Torrance, CA, USA). The column temperature was maintained at 25 °C. Eluent A was water and eluent B was MeOH, both containing 0.1 % formic acid. The chromatographic method held the initial mobile phase composition (10 % B) constant for 2 min, followed by a linear gradient to 100 % B within 30 min. This final condition was held for 5 min, followed by 8 min of column re-equilibration at 10 % B. The flow rate of the mobile phase was 250 $\mu L\,min^{-1}$ and the injection volume was 10 μL.

The ESI interface was used in positive ion mode with the following settings: sheath gas, 60 arbitrary units; auxiliary gas, 15 arbitrary units; sweep gas, 5 arbitrary units; capillary voltage, 4 kV; capillary temperature, 300 °C. All other source parameters were automatically tuned for maximum signal intensity of a 10-$mg\,L^{-1}$ reserpine solution (Sigma-Aldrich). For the FT-Orbitrap, the automatic gain control was set to a target value of 3×10^5, and a maximum injection time of 500 ms was chosen. The mass spectrometer was operated in a scan range from *m*/*z* 100 to *m*/*z* 1,000 with a resolving power setting of 60,000 FWHM (at *m*/*z* 400). Data were recorded using Xcalibur 2.1.0 (Thermo Fisher Scientific). For MS/MS measurements, collision-induced dissociation (Q=0.250, activation time 30 ms, resolving power setting 15,000 FWHM) was used.

Data processing and annotation of DON conjugates

An improved yet unpublished version of MetExtract [12], developed by us, was applied to automatically extract corresponding MS peak pairs in mass spectra of a 1+1 mixture of biological samples containing natural and ^{13}C fully labelled xenobiotics. Putative DON-metabolisation products had to fulfil the following criteria: (1) the monoisotopic non-labelled and the completely labelled isotopologues of DON-derivative ions form the principal ions of their corresponding isotopic patterns and have to be present in at least two mass spectra, (2) peak area ratio of monoisotopic peaks and corresponding fully labelled analogues has to be 0.8–1.2 and (3) monoisotopic non-labelled and the completely labelled isotopologue ions have to show chromatographic co-elution.

Furthermore, within all LC-HRMS chromatograms, putative DON-derived ion signals were grouped according to retention time with the aim to deconvolute mass spectra and evaluate the number of different metabolites as well as type of ion species.

Different metabolites were characterised and annotated according to following criteria: (1) the accurate mass differs less than ±5 ppm from the theoretical postulated mass and (2) possible heteroatoms within the conjugates (e.g. sulphur) are determined by evaluating the isotopic pattern of the completely labelled isotopologue. Moreover, MS/MS spectra were recorded and evaluated using Thermo Xcalibur 2.1.0.

Results and discussion

Detection and characterisation of DON-biotransformation products

MetExtract data processing revealed a total of 57 ion pairs in full-scan chromatograms of DON-treated samples that showed the monoisotopic non-labelled and the corresponding completely labelled isotopologue of the DON moiety. No ion pairs could be detected in mock-treated samples. The obtained features were sorted according to retention time and grouped into ten distinct chromatographic peak groups with the aid of MetExtract. DON (*m*/*z* 297.1327, $[M+H]^+$) and nine DON-biotransformation products were detected in wheat samples treated with a 1+1 mixture of native and $^{13}C_{15}$-labelled DON. The corresponding peak pairs of all detected DON-biotransformation products showed a mass difference of exactly 15.0503 Da indicating the presence of all 15 carbon atoms of DON. Thus, these DON metabolites contained the intact carbon skeleton of DON as part of their molecular structure. No DON degradation products with less than 15 carbon atoms were found. Moreover, all determined masses were different and clearly higher than the molecular weight of DON, indicating the formation of several structurally different DON conjugates by wheat. Figure 1 depicts the extracted ion currents of the non-labelled and ^{13}C-labelled metabolite ions representing all 57 features of the detected compounds.

Among the detected substances, the identity of DON (peak #6) and D3G (*m*/*z* 459.1848, $[M+H]^+$, peak #7) was confirmed by standard measurements. This result shows that also the susceptible cultivar Remus has the capacity to form D3G [15].

A closer manual inspection of all isotopic patterns of both the ^{12}C monoisotopic as well as the fully ^{13}C-labelled ions

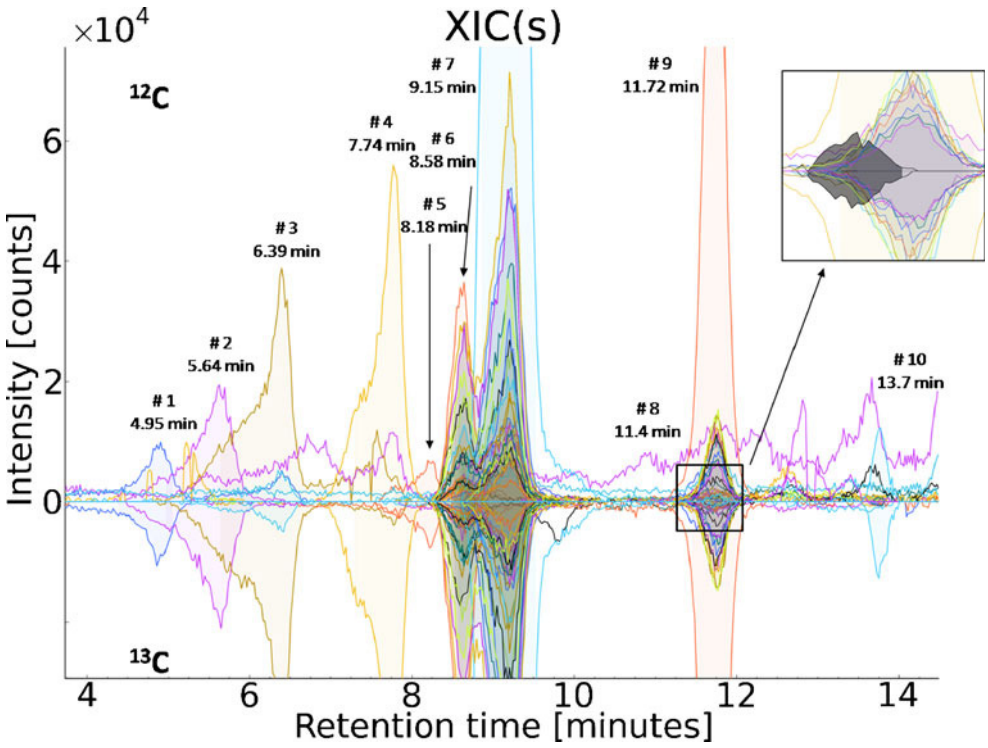


Fig. 1 EICs extracted by MetExtract: positive intensities show the EICs of the monoisotopic ^{12}C labelled metabolite ions, while negative intensities represent the EICs of the corresponding ^{13}C fully labelled metabolite ions. All of the *m/z* values labelled with retention time and number showed similar peak area, peak shape and perfect co-elution and thus can be regarded as signals originating from biotransformation products of DON (*#2* DON-S-Cys, *#3* DON-S-Cys-Gly, *#4* DON-GSH, *#6* DON, *#7* D3G)

revealed the presence of sulphur in three of the detected DON conjugates. The isotope ^{34}S was confirmed by the presence of a mass peak with a relative intensity of about 4 % and a mass difference of +1.9956 Da from the principal ions in the respective spectra. Most importantly, our study identified DON-GSH (*m/z* 604.2158, $[M+H]^+$, peak #4) to be formed by wheat. $[M+Na]^+$ ions for DON-GSH were observed; the molecular mass of the intact metabolite was determined with a mass error of −2.4 ppm. Additionally, two processing products of the glutathione conjugate of DON were found, peak #2 is putatively identified as DON-*S*-cysteine (−3.2 ppm) and peak #3 was annotated as DON-*S*-cysteinyl-glycine (−2.2 ppm).

In addition to D3G and the glutathione-related conjugates, five currently unknown DON-biotransformation products (peaks #1, #5, #8, #9 and #10) were found. For a more detailed structural characterisation of the detected DON-GSH, MS/MS measurements using collision-induced dissociation were carried out.

Characterisation of DON-GSH, DON-*S*-cysteinyl-glycine and DON-*S*-cysteine by LC-MS/MS measurements

Further, MS/MS measurements of DON-treated wheat using the precursor mass *m/z* 604.21 (corresponding to $[DON-GSH+H]^+$) at 18 eV revealed the characteristic fragmentation behaviour of GSH adducts described by Levsen et al. [20]. Losses of glycine (75 Da), anhydroglutamic acid (129 Da), glutamine (146 Da), as well as the cleavage of the S-CH_2 bond (275 Da) and DON (296 Da) for DON-GSH were observed (Fig. 2a). For structure confirmation of DON-*S*-Cys-Gly and DON-*S*-Cys, further MS/MS measurements of the precursor ions *m/z* 475.18 (at 25 eV) and *m/z* 418.15 (at 22 eV) have been carried out respectively (Fig. 2b, c). For both DON conjugates, the MS/MS spectra contained an abundant loss of NH_3 (17 Da), the cleavage of the S-CH_2 bond and $[DON+H]^+$ which is in good agreement with the characteristic MS/MS fragments of GSH conjugates as described earlier [20].

Previously, only D3G has been reported in the literature as detoxification product of DON *in planta* [21]. Based on the transcriptome response of barley after DON treatment, the formation of DON-GSH conjugates has been postulated, and the non-enzymatic formation of a complex mixture of glutathione conjugates has been demonstrated by NMR [22]. To our best knowledge, our work reports the formation of DON-GSH *in planta* for the first time. The addition of GSH to the double bond of DON, for which NMR evidence was provided [22], seems most likely and therefore this structure (without consideration of its stereochemistry) is shown in Fig. 2. The formation of the glutathione adducts with α,β-unsaturated ketones is a reversible reaction [23], effectively preventing their isolation. Moreover, the occurrence of the glutathione conjugate processing products DON-*S*-Cys-Gly and the DON-*S*-Cys conjugate in wheat are described here for the first time. Our data are in good agreement with well-known detoxification reactions of xenobiotics such as herbicides *in planta* via GSH

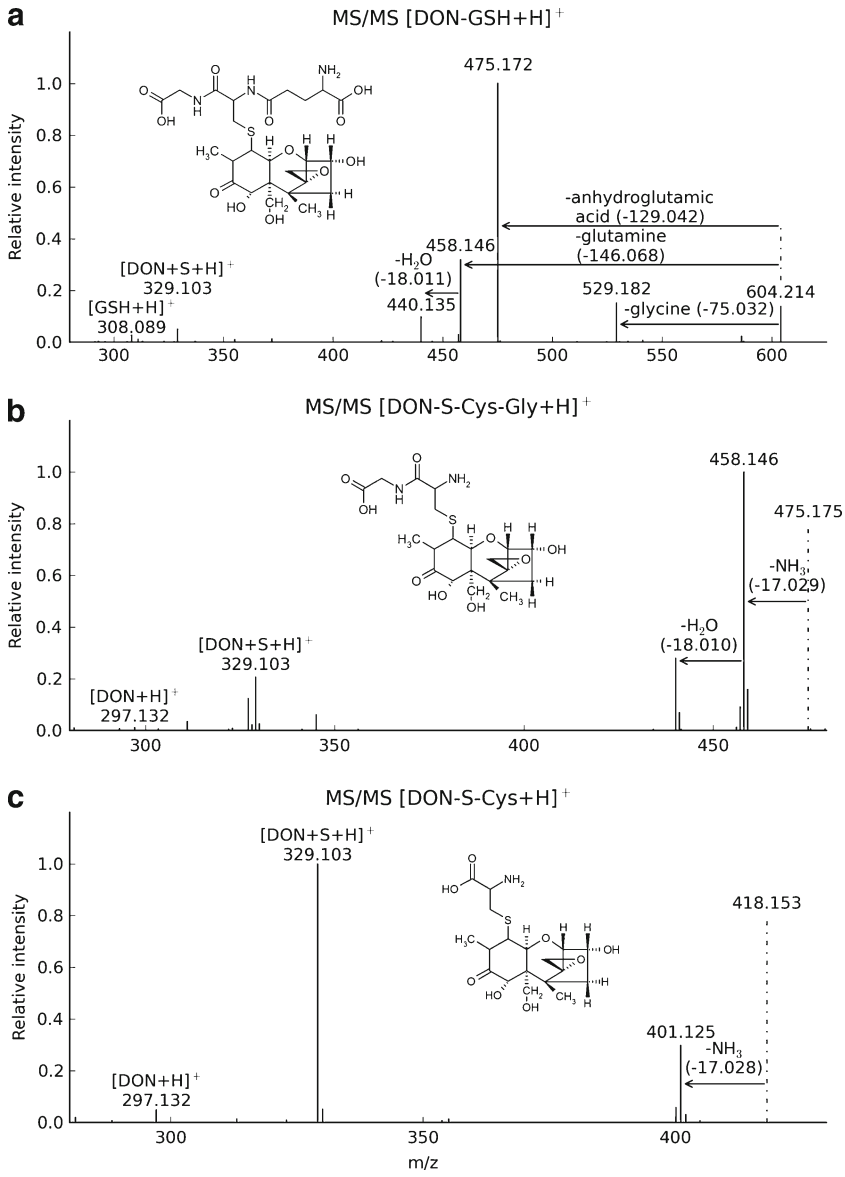


Fig. 2 CID MS/MS spectra of DON-GSH (precursor ion, *m/z* 604.21), DON-S-Cys-Gly (precursor ion, *m/z* 475.18) and DON-S-Cys (precursor ion, *m/z* 418.15) with suggested structure formula

conjugation and subsequent hydrolysis of the formed conjugates in the vacuole of plant cells [24]. Furthermore, our results do also fit to the observation in barley, where DON treatment triggered a strong upregulation of cysteine biosynthesis suggesting that cysteine is used for glutathione formation and conjugation to DON [22]. The relevance of DON-GSH and its GSH-related biotransformation products regarding food safety is currently unknown and has to be investigated in future studies. Furthermore, additional studies will be needed to elucidate the structure of the remaining, currently unknown DON conjugates. These preliminary results show the high potential of the presented approach, which can be used to study the metabolisation of all types of xenobiotics independent of the nature of the biological system.

Conclusion

The presented SIL-assisted metabolomics approach in combination with LC-HRMS is a powerful tool for the untargeted screening of biotransformation products of xenobiotics in virtually all types of biological systems. Labelling-specific isotopic patterns can be reliably and automatically detected after measurement of biological samples treated with a 1+1 mixture of labelled and non-labelled precursors. Moreover, the data patterns directly reveal the number of carbon atoms in the detected metabolite ions. In this preliminary study, the great potential of the presented approach is further underlined by the successful and automated detection of eight novel plant-derived biotransformation products of the mycotoxin

DON. The detection of the DON-GSH conjugate and derived processing products in wheat is reported for the first time, providing evidence for glutathione-mediated metabolism and presumably detoxification of the *Fusarium* virulence factor DON *in planta*.

Acknowledgments The authors greatly acknowledge the Austrian Science Fund (project SFB Fusarium 3706-B11) for funding this research and the Federal Country Lower Austria as well as the European Regional Development Fund of the European Union (grant number GZ WST3-T-95/001-2006) for financial support.

References

1. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol 22(5):245–252
2. Patterson AD, Gonzalez FJ, Idle JR (2010) Xenobiotic metabolism: a view through the metabolometer. Chem Res Toxicol 23 (5):851–860. doi:10.1021/tx100020p
3. Chen C, Gonzalez FJ, Idle JR (2007) LC-MS-based metabolomics in drug metabolism. Drug Metab Rev 39(2–3):581–597. doi:10.1080/03602530701497804
4. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13(4):263–269
5. Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J (2005) Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends Biotechnol 23(1):28–33
6. Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H, Harms AC, Ulrich EL, Markley JL, Sussman MR (2007) Stable isotope assisted assignment of elemental compositions for metabolomics. Anal Chem 79(18):6912–6921. doi:10.1021/ac070346t
7. Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L (2008) High-resolution direct infusion-based mass spectrometry in combination with whole 13c metabolome isotope labeling allows unambiguous assignment of chemical sum formulas. Anal Chem 80(24):9417–9425. doi:10.1021/ac8014627
8. Giavalisco P, Köhl K, Hummel J, Seiwert B, Willmitzer L (2009) 13C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography–mass spectrometry-based metabolomic research. Anal Chem 81 (15):6546–6551. doi:10.1021/ac900979e
9. Giavalisco P, Li Y, Matthes A, Eckhardt A, Hubberten H-M, Hesse H, Segu S, Hummel J, Köhl K, Willmitzer L (2011) Elemental formula annotation of polar and lipophilic metabolites using 13C, 15 N and 34S isotope labelling, in combination with high-resolution mass spectrometry. Plant J 68(2):364–376. doi:10.1111/j.1365-313X.2011.04682.x
10. Hiller K, Metallo C, Stephanopoulos G (2011) Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics. Curr Pharm Biotechnol 12(7):1075–1086
11. Athersuch TJ, Nicholson JK, Wilson ID (2007) Isotopic enrichment enhancement in metabonomic analysis of UPLC-MS data sets. J Label Compd Radiopharm 50(5–6):303–307. doi:10.1002/jlcr.1217
12. Bueschl C, Kluger B, Berthiller F, Lirk G, Winkler S, Krska R, Schuhmacher R (2012) MetExtract: a new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. Bioinformatics. doi:10.1093/bioinformatics/bts012
13. Pestka J (2010) Deoxynivalenol: mechanisms of action, human exposure, and toxicological relevance. Arch Toxicol 84(9):663–679. doi:10.1007/s00204-010-0579-8
14. Walter S, Nicholson P, Doohan FM (2010) Action and reaction of host and pathogen during *Fusarium* head blight disease. New Phytol 185(1):54–66. doi:10.1111/j.1469-8137.2009.03041.x
15. Lemmens M, Scholz U, Berthiller F, Dall'Asta C, Koutnik A, Schuhmacher R, Adam G, Buerstmayr H, Mesterházy Á, Krska R, Ruckenbauer P (2005) The ability to detoxify the mycotoxin deoxynivalenol colocalizes with a major quantitative trait locus for fusarium head blight resistance in wheat. Mol Plant-Microbe Interact 18(12):1318–1324. doi:10.1094/mpmi-18-1318
16. Proctor RH, Hohn TM, McCormick SP (1995) Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Mol Plant-Microbe Interact 8 (4):593–601
17. Poppenberger B, Berthiller F, Lucyshyn D, Sieberer T, Schuhmacher R, Krska R, Kuchler K, Glössl J, Luschnig C, Adam G (2003) Detoxification of the fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from *Arabidopsis thaliana*. J Biol Chem 278(48):47905–47914. doi:10.1074/jbc.M307552200
18. Berthiller F, Dall'Asta C, Schuhmacher R, Lemmens M, Adam G, Krska R (2005) Masked mycotoxins: determination of a deoxynivalenol glucoside in artificially and naturally contaminated wheat by liquid chromatography–tandem mass spectrometry. J Agric Food Chem 53(9):3421–3425. doi:10.1021/jf047798g
19. De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protocols 2 (4):778–791
20. Levsen K, Schiebel H-M, Behnke B, Dötzer R, Dreher W, Elend M, Thiele H (2005) Structure elucidation of phase II metabolites by tandem mass spectrometry: an overview. J Chromatogr A 1067 (1–2):55–72
21. Berthiller F, Schuhmacher R, Adam G, Krska R (2009) Formation, determination and significance of masked and other conjugated mycotoxins. Anal Bioanal Chem 395(5):1243–1252. doi:10.1007/s00216-009-2874-x
22. Gardiner SA, Boddu J, Berthiller F, Hametner C, Stupar RM, Adam G, Muehlbauer GJ (2010) Transcriptome analysis of the barley–deoxynivalenol interaction: evidence for a role of glutathione in deoxynivalenol detoxification. Mol Plant-Microbe Interact 23(7):962–976. doi:10.1094/mpmi-23-7-0962
23. Schwöbel JAH, Madden JC, Cronin MTD (2010) Examination of Michael addition reactivity towards glutathione by transition-state calculations. SAR QSAR Environ Res 21(7–8):693–710. doi:10.1080/1062936x.2010.528943
24. Marrs KA (1996) The functions and regulation of glutathione s-transferases in plants. Annu Rev Plant Physiol Plant Mol Biol 47 (1):127–158. doi:10.1146/annurev.arplant.47.1.127

Publication #3: A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research

Christoph Bueschl†, Bernhard Kluger†, Marc Lemmens, Gerhard Adam, Gerlinde Wiesenberger, Valentina Maschietto, Adriano Marocco, Joseph Strauss, Stephan Bödi, Gerhard G. Thallinger, Rudolf Krska, Rainer Schuhmacher
Metabolomics. 2014;10:754-769

Contributions of the presenting author: The author of this thesis significantly extended the software MetExtract. This updated version of the tool was used to perform data processing of the *F. graminearum*, wheat and maize derived LC-HRMS data. These results were used to demonstrate the developed workflow and the benefits of full metabolome labelling applications in untargeted metabolomics research. Furthermore, the author of this thesis performed statistical evaluation of the generated data to show the benefits of metabolome-wide internal standardisation and together with Bernhard Kluger wrote the publication. Since this publication heavily depends on the interdisciplinary research between analytical chemistry and bioinformatics, the authors agreed that this publication was realised with shared first author option.

† Authors have contributed equally to this work

Metabolomics (2014) 10:754–769
DOI 10.1007/s11306-013-0611-0

ORIGINAL ARTICLE

A novel stable isotope labelling assisted workflow for improved untargeted LC–HRMS based metabolomics research

Christoph Bueschl · Bernhard Kluger · Marc Lemmens · Gerhard Adam · Gerlinde Wiesenberger · Valentina Maschietto · Adriano Marocco · Joseph Strauss · Stephan Bödi · Gerhard G. Thallinger · Rudolf Krska · Rainer Schuhmacher

Received: 25 October 2013 / Accepted: 26 November 2013 / Published online: 4 December 2013

Abstract Many untargeted LC–ESI–HRMS based metabolomics studies are still hampered by the large proportion of non-biological sample derived signals included in the generated raw data. Here, a novel, powerful stable isotope labelling (SIL)-based metabolomics workflow is presented, which facilitates global metabolome extraction, improved metabolite annotation and metabolome wide internal standardisation (IS). The general concept is exemplified with two different cultivation variants, (1) co-cultivation of the plant pathogenic fungi *Fusarium graminearum* on non-labelled and highly ^{13}C enriched culture medium and (2) experimental cultivation under native conditions and use of globally U-^{13}C labelled biological reference samples as exemplified with maize and wheat. Subsequent to LC–HRMS analysis of mixtures of labelled and non-labelled samples, two-dimensional data filtering of SIL specific isotopic patterns is performed to better extract truly biological derived signals together with the corresponding number of carbon atoms of each metabolite ion. Finally, feature pairs are convoluted to feature groups each representing a single metabolite. Moreover, the correction of unequal matrix effects in different sample types and the improvement of relative metabolite quantification with metabolome wide IS are demonstrated for the *F. graminearum* experiment. Data processing employing the presented workflow revealed about 300 SIL derived feature pairs corresponding to 87–135 metabolites in *F. graminearum* samples and around 800 feature pairs corresponding to roughly 350 metabolites in wheat samples. SIL assisted IS, by the use of globally U-^{13}C labelled biological samples, reduced the median CV value from 7.1 to 3.6 % for technical replicates and from 15.1 to 10.8 % for biological replicates in the respective *F. graminearum* samples.

Keywords ^{13}C-labelling · Internal standardisation · Metabolomics · Fusarium · Wheat · Maize

Christoph Bueschl and Bernhard Kluger have contributed equally to this work.

C. Bueschl · B. Kluger · M. Lemmens · R. Krska · R. Schuhmacher (✉)
Department for Agrobiotechnology (IFA-Tulln), Center for Analytical Chemistry and Institute for Biotechnology in Plant Production, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Str. 20, 3430 Tulln, Austria
e-mail: rainer.schuhmacher@boku.ac.at

G. Adam · G. Wiesenberger · J. Strauss · S. Bödi
Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Str. 24, 3430 Tulln, Austria

V. Maschietto · A. Marocco
Institute of Agronomy, Genetics and Field Crops, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy

J. Strauss
Health and Environment Department, Bioresources – Fungal Genetics and Genomics, Austrian Institute of Technology (AIT), Konrad-Lorenz-Str. 24, 3430 Tulln, Austria

G. G. Thallinger
Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

G. G. Thallinger
Core Facility Bioinformatics, Austrian Centre for Industrial Biotechnology, Petersgasse 14, 8010 Graz, Austria

1 Introduction

While full genome sequences have been determined for many organisms, it is currently still not possible to measure the complete metabolite inventory of a biological system due to methodical limitations. Complementary, sensitive and generic techniques are required to cope with the large chemical diversity and wide dynamic range of low molecular weight metabolites. Gas chromatography (GC) or liquid chromatography (LC) coupled to mass spectrometry (MS) as well as nuclear magnetic resonance (NMR) spectroscopy have emerged as key techniques in the field of metabolomics, as recently reviewed by e.g. Zhang et al. (2012), Patti et al. (2012b) and Zhou et al. (2012). The combination of LC with electrospray ionisation (ESI) high resolution mass spectrometry (HRMS) has proven to be particularly powerful as this technique enables the detection of a large number of known and unknown metabolites simultaneously and requires only small amounts of the biological sample (Hiller et al. 2011; Patti et al. 2012b).

Two different metabolomics concepts can be distinguished: targeted and untargeted approaches. In targeted approaches, a set of predefined known substances is determined, thus, absolute quantification of those metabolites, which are available as authentic reference standards, is feasible. In contrast, untargeted approaches try to find mass spectrometric features of all detectable metabolites, including those unknown or at least unidentified at the time of measurement. Therefore, the untargeted approach has the advantage of probing the entire, observable metabolic space and can obtain relative abundances of several hundreds to thousands of metabolites simultaneously (Patti et al. 2012b). For the automated data processing of such LC–HRMS derived metabolomics datasets, various workflows and software packages have been developed and are frequently used in untargeted metabolomics studies e.g. XCMS (Smith et al. 2006), MzMine (Pluskal et al. 2010), MetAlign (Lommen and Kools 2012) or Maven (Clasquin et al. 2012). These software tools have in common, that they extract as many features as possible from raw LC–HRMS derived metabolomics data sets. In this respect the term feature has been defined to be a bounded, two dimensional LC–HRMS signal consisting of a chromatographic peak (i.e. retention time) and a MS signal (*m/z* value) (Kuhl et al. 2012).

Despite the recent advances regarding both LC–HRMS instrumentation and data handling platforms, the comprehensive annotation of the metabolome of a biological sample of interest and subsequent metabolite identification still remain the major bottlenecks in untargeted metabolomics, especially for LC-ESI-HRMS based studies (Scalbert et al. 2009; Castillo et al. 2011; Patti et al. 2012b; Theodoridis et al. 2012; Dunn et al. 2013). This limitation can largely be attributed to the generic nature of the ESI process, unavoidably leading to LC-ESI-HRMS full scan chromatograms and spectra, containing a large proportion of background and chemical noise compared to the signals originating from true metabolites (Keller et al. 2008; Covey et al. 2009; Trotzmüller et al. 2011). Further challenges arise from the fact that a single metabolite leads to more than one ion species (e.g. isotopologue peaks, different adducts, in-source fragments and even more complex combinations of the previous species). In addition, many metabolites cannot completely be separated in the chromatographic dimension and therefore LC–HRMS measurements result in mass spectra, which contain signals from more than one metabolite.

Another obstacle of untargeted LC-ESI-HRMS based metabolomics is related to relative quantification of the detected metabolite ions, which is caused by so called matrix effects. The composition of the evaporated sample at any time point of the LC–HRMS measurement can significantly influence the ionization efficiency and leads to ion suppression or ion enhancement in the ESI source of the mass spectrometer (Tang and Kebarle 1993; King et al. 2000). Matrix effects can seriously affect signal intensities as well as precision and even limit the coverage of the metabolome (Vogeser and Seger 2010; Koal and Deigner 2010). They are difficult to overcome in global untargeted studies as the matrix is composed of the biological sample itself. Thus, except protein precipitation, sample purification is generally not a suitable option as this would largely discriminate many sample constituents of interest (Tulipani et al. 2013). Moreover, the availability of appropriate internal standards is often limited. The detailed and comprehensive study of matrix effects is laborious and challenging, thus only a few studies reported the systematic evaluation of matrix effects and their limitations on relative metabolite quantification in the field of LC–HRMS based metabolomics (Böttcher et al. 2007; Redestig et al. 2011; Tulipani et al. 2013).

With respect to the above mentioned limitations regarding global annotation of the metabolome and method performance evaluation, there is a great demand for both innovative approaches for the analytical measurement of biological samples with LC–HRMS as well as the development of novel, improved data processing algorithms.

Stable isotope labelling (SIL) is a technique, which is becoming increasingly used in different areas of metabolomics research and it shows the potential to conquer many of the elucidated limitations in untargeted metabolomics research. In this respect, SIL assisted experiments employ stable isotopes of elements such as carbon (^{13}C), hydrogen (^{2}H), oxygen (^{18}O), nitrogen (^{15}N) and sulphur (^{34}S) (Klein and Heinzle 2012; Nakabayashi et al. 2013)

respectively. However, ^{13}C is used most commonly as the main labelling isotope, since carbon is part of virtually any metabolite. Non-labelled, partly labelled and highly (>98 %) ^{13}C enriched (U-^{13}C) metabolites show the same physico-chemical properties and therefore are not separated by chromatography, but can easily be distinguished by their mass to charge ratio (*m/z*) using an MS instrument.

The use of globally U-^{13}C labelled biological samples enables to circumvent problems in untargeted metabolomics, such as metabolome annotation, generation of sum formulas of the detected metabolites and putative metabolite identification. It was demonstrated that the combination of ^{13}C, ^{15}N and ^{34}S labelling for example can help to assign the number of atoms of the respective labelling element to a metabolite ion correctly and thereby facilitates annotation of metabolites by database search (Hegeman et al. 2007; Giavalisco et al. 2008; Cano et al. 2013). In addition to improved feature extraction and metabolite annotation, SIL experiments have also been successfully used to accomplish internal standardisation (IS) for quantification of metabolite levels and, thus correct ion suppression or MS signal fluctuations caused by matrix effects in LC-ESI-HRMS (Bennett et al. 2008; Giavalisco et al. 2009; Hegeman 2010). Moreover, IS by globally stable isotope labelled biological samples allow both detailed characterisation of the performance of the used metabolomics workflow as well as an improved relative quantification/technical precision of metabolomics data.

Despite the high potential of SIL assisted approaches and their successful application in various fields, to the best of our knowledge only a few data processing tools have been published for the automated evaluation of LC–HRMS data originating from labelled biological samples. For non-targeted GC–MS based metabolomics SIL assisted metabolomics, Hiller and colleagues published the Non-targeted Tracer Fate Detection (NTFD) algorithm (Hiller et al. 2013) to study labelled tracer compounds in the central metabolism (Hiller et al. 2010). Moreover, de Jong and Beecher (2012) have successfully implemented a sophisticated method termed IROA (Isotopic Ratio Outlier Analysis^{TM}) to automatically extract features differing between experimental conditions after parallel cultivation on native and U-^{13}C-labelled nutrition sources. IROA is offered as a commercial metabolomics application and software programme. The R package mzMatch-ISO (Chokkathukalam et al. 2013) is a software tool for the annotation and relative quantification of SIL derived MS data with the aim to provide insight into metabolic fluxes of biological systems. It is designed to use metabolomics data analysed by XCMS (Tautenhahn et al. 2008) and allows an in depth evaluation and visualisation of the associated isotopic patterns and their respective abundances of various native and labelled metabolites. To the best of our knowledge, MetExtract, which has been developed in our laboratory, is to date the only publicly available tool aiming at the untargeted, automated global detection of truly metabolite derived LC–HRMS signals originating from natural (compounds showing a natural carbon isotopic distribution pattern are termed "non-labelled" in the following) and stable isotope labelled biological samples (Bueschl et al. 2012).

Here a detailed analytical and data processing workflow for SIL assisted untargeted LC–HRMS based metabolomics experiments is presented. This workflow is exemplified by two representative experiments: In the first approach the filamentous fungus *Fusarium graminearum* is cultivated in parallel on a non-labelled and a U-^{13}C labelled carbon source respectively under identical conditions. In the second approach a metabolomics experiment is performed using a non-labelled carbon source for cultivation of biological samples (wheat and maize), while globally U-^{13}C labelled biological reference samples are used for IS. The concept and performance of both variants are presented in detail.

2 Materials and methods

2.1 Chemicals and biological samples

Acetonitrile (ACN, HiPerSolv Chromanorm, HPLC gradient grade) was purchased from VWR (Vienna, Austria); Methanol (MeOH, LiChrosolv, LC gradient grade) was purchased from Merck (Darmstadt, Germany); formic acid (FA, MS grade) was obtained from Sigma-Aldrich (Vienna, Austria). Water was purified successively by reverse osmosis and an ELGA Purelab Ultra-AN-MK2 system (Veolia Water, Vienna, Austria). Components of the modified FMM were purchased from the following suppliers: Fluka (KH_2PO_4, $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$), Roth ($MgSO_4{\cdot}7H_2O$, KCl, $ZnSO_4$, H_3BO_3), Sigma Aldrich ($NaNO_3$, $MnSO_4$, $CuSO_4{\cdot}5H_2O$, $Na_2MoO_4{\cdot}2H_2O$), Serva (citric acid) and VWR (glucose). U-$^{13}C_6$-glucose with a ^{13}C enrichment degree of 99 % was obtained from Eurisotop (Saarbrücken, Germany). U-^{13}C labelled wheat ear (>97 % ^{13}C, cultivar Baldus), and U-^{13}C labelled maize kernels (>97 % ^{13}C, cultivar Yukon chief) were obtained from Isolife (Wageningen, The Netherlands).

2.2 Cultivation of *Fusarium graminearum*, wheat and maize samples

In this study a metabolomics workflow for two different cultivation variants using stable isotope labelling is presented (Fig. 1). In the first approach (thereafter referred to as variant A) the biological organism of interest is co-cultivated under identical conditions, using either a ^{12}C or

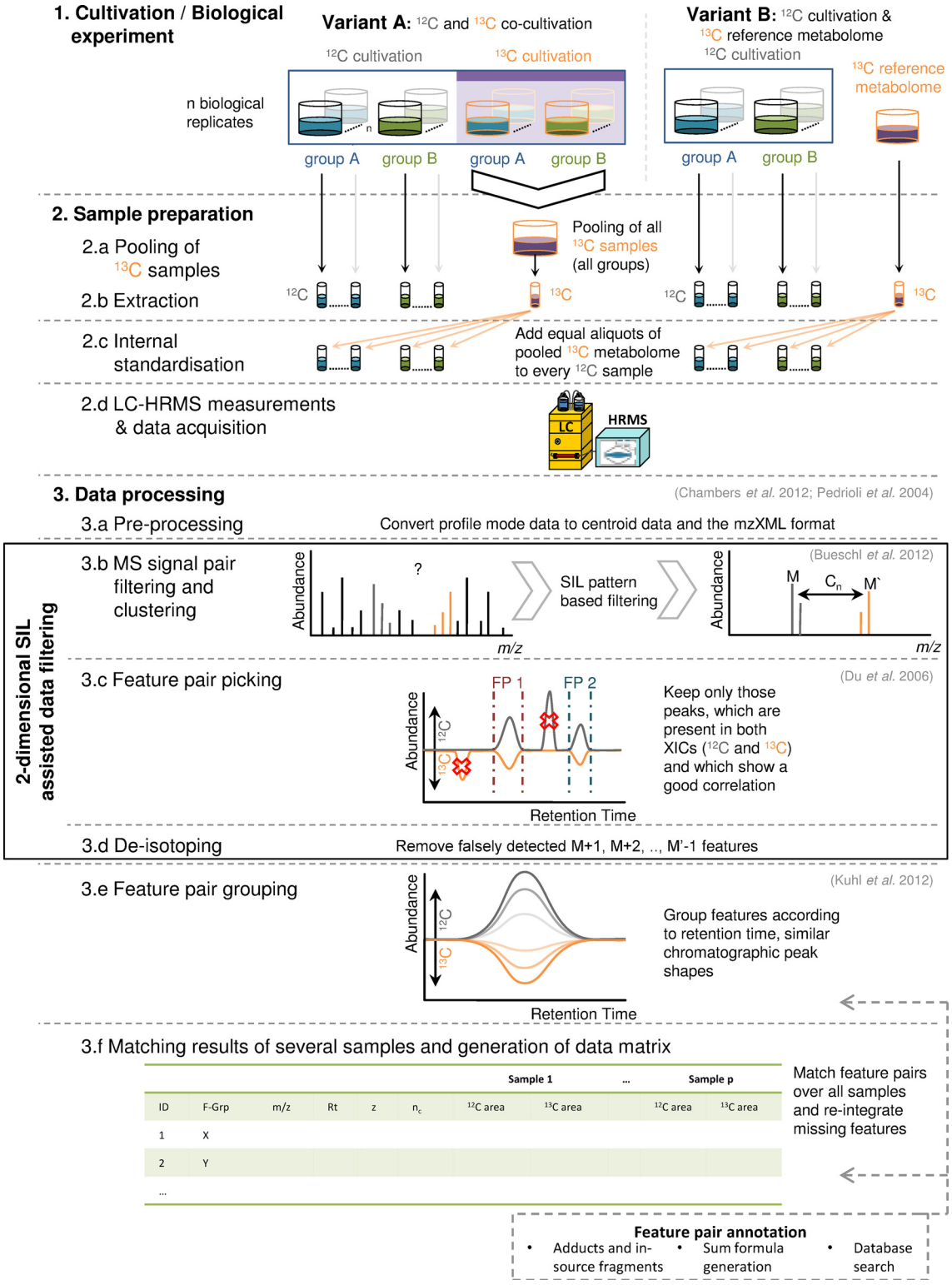


Fig. 1 Overview of the proposed SIL assisted workflow for native and U-^{13}C co-cultivation (variant A) and native cultivation and use of U-^{13}C reference metabolome (variant B) [figure-width: 174 mm]

^{13}C carbon source respectively. This approach is favoured for less complex organisms such as bacteria, yeasts or filamentous fungi which allow cultivation on a defined minimal medium, where the natural carbon source can be easily replaced by highly (>98 %) ^{13}C- or ^{15}N-enriched nutrients. Variant A is demonstrated for the filamentous

fungus *F. graminearum*, a pathogen of several cereal crops. In the second approach (referred to as variant B) a metabolomics experiment is performed using a native ^{12}C carbon source for cultivation, while a globally U-^{13}C labelled biological sample serves as reference metabolome for IS of the non-labelled experimental samples. This approach is preferred when isotope labelling under experimental conditions is difficult to achieve or not feasible, e.g. with animals or plants. Experimental details are given to a degree necessary to fit the purpose of this paper, which is to present and discuss the analytical concept and data processing rather than the whole biological study.

2.2.1 Variant A: ^{12}C and ^{13}C co-cultivation (*F. graminearum* samples)

Fusarium graminearum wild-type (PH-1, NRRL 31084) and its isogenic *tri5*Δ::*loxP* mutant lacking the gene encoding trichodiene synthase, the first enzyme in the trichothecene biosynthetic pathway were used. The strains were cultivated in a modified Fusarium minimal medium (FMM: 1 g/L KH_2PO_4, 0.5 g/L $MgSO_4{\cdot}7H_2O$, 0.5 g/L KCl, 2 g/L $NaNO_3$, 10 mg/L citric acid, 10 mg/L $ZnSO_4{\cdot}6H_2O$, 2 mg/L $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$, 0.5 mg/L $CuSO_4{\cdot}5H_2O$, 0.1 mg/L $MnSO_4$, 0.1 mg/L H_3BO_3, 0.1 mg/L $Na_2MoO_4{\cdot}2H_2O$) (Leslie and Summerell 2007) containing either non-labelled glucose or U-^{13}C-glucose as sole carbon source at a concentration of 10 g/L. *F. graminearum* wild-type and *tri5*Δ strains were grown on non-labelled as well as on U-^{13}C labelled FMM using six biological replicates per strain and nutrition condition resulting in a total of 24 samples. The cultures were set up as follows: the strains were sporulated in mung-bean medium as described before by Kluger et al. (2013). 1-ml aliquots of either non-labelled or U-^{13}C labelled glucose containing medium were pipetted to each well of a UNIFILTER 24-well 10 ml filtration microplate equipped with a Whatman GF/C filter (VWR, Vienna, Austria) and each well was inoculated with 2,000 spores of the respective *F. graminearum* strain. Still cultures were grown at 20 °C in the dark for 7 days.

2.2.2 Variant B: ^{12}C cultivation & ^{13}C reference metabolome (wheat and maize samples)

Seeds of the wheat cultivars "Remus" and "CM-82036" were grown in pots with soil under environmentally controlled greenhouse conditions. Light and watering regime, humidity and temperature were kept under controlled conditions whenever possible and readjusted continuously to fit the plant's actual developmental stage. At the onset of anthesis five ears (ten spikelets per ear) were harvested for each cultivar at the beginning of a luminescence cycle and immediately shock-frozen in liquid nitrogen to quench cellular metabolism. In total ten ears were sampled and stored at −80 °C until further sample preparation.

Seeds of the maize line CO354 were planted in pots and later transferred to an environmentally controlled greenhouse with controlled light regime and temperature conditions. Ears were harvested 18 days after hand pollination and kernels were immediately extracted using sterilised scalpels. Three pools of kernels were obtained, where each pool derived from the mixing of seeds came from three different ears, samples were immediately frozen in liquid nitrogen after collecting and stored at −80 °C for further analyses.

For IS using a U-^{13}C labelled reference metabolome according to variant B (Fig. 1), a U-^{13}C labelled wheat ear (>97 % ^{13}C, cultivar Baldus), and U-^{13}C labelled maize kernels (>97 % ^{13}C, cultivar Yukon chief) were used respectively. It was taken into consideration that the labelled plant material had been grown to the same development stage as the non-labelled wheat cultivars and the native maize line CO354 respectively.

2.3 Sample preparation

2.3.1 Variant A: preparation of *F. graminearum* samples

The 24-well microtiter plate was removed from the climate chamber and immediately centrifuged for 10 min at 2,000 rpm to separate the mycelium from the extracellular metabolites in the supernatant. Non-labelled and U-^{13}C labelled supernatants were prepared in parallel according to the following protocol. 500 µl aliquots of supernatants of wildtype and *tri5*Δ mutant, which had been grown on native glucose containing FMM were transferred separately each into a 1.5 ml Eppendorf tube resulting in a total of 12 samples ($n = 6$ replicates per fungal strain). 400 µl aliquots of each of U-^{13}C labelled supernatants were pooled together in a 50 ml polystyrene tube (VWR International GmbH, Vienna, Austria) resulting in a total of 4,800 µl of a pooled U-^{13}C-supernatant for IS. Immediately after centrifugation and aliquoting of supernatants, all aliquots were quenched with 30 % acetonitrile (v/v) resulting in a 7:3 (v/v) ratio of supernatant to acetonitrile. For LC–HRMS analysis, each quenched non-labelled supernatant was standardised by adding the same volume (200 µl) of the pooled and quenched ^{13}C-supernatant resulting in LC–HRMS sample aliquots (1:1, v/v).

In addition, 60 µl of each of the U-^{13}C-standardised LC–HRMS samples ($n = 6$ replicates per fungal strain) were merged to an aggregate sample (AG) which was used to evaluate the precision of the LC–HRMS measurement and chromatographic peak integration steps by repeated injection out of the same HPLC vial ($n = 13$ replicates).

LC–HRMS analysis of all samples was carried out immediately after sample preparation.

For comparison of SIL assisted data processing and conventional data processing by XCMS as well as to demonstrate the extraction efficiency, aliquots of quenched native supernatants without any U-^{13}C labelled material were mixed with quenched FMM (1:1, v/v) to yield LC–HRMS samples exhibiting same concentration levels of non-labelled metabolites as the U-^{13}C labelled standardised analogues.

To further exemplify the selectivity of the proposed workflow to identify only SIL derived biological information, solvent blanks (water:acetonitrile (7:3, v/v)) containing purified water instead of supernatant were prepared in parallel according to the same procedure mentioned above.

2.3.2 *Variant B: preparation of wheat and maize samples*

The sample preparation of plant material was based on De Vos et al. (2007) carried out after slight modifications as reported in Kluger et al. (2012). Native wheat ears and U-^{13}C labelled wheat ear "Baldus" were extracted and prepared in parallel. Native wheat ears of the cultivar "Remus" and "CM-82036" respectively ($n = 5$ replicates per cultivar) were milled separately to a fine powder using a ball mill (MM301 Retsch, Haan, Germany). 100 ± 5 mg of homogenised plant material were weighed to 1.5 mL-Eppendorf tubes with subsequent extraction using 1 mL of pre-cooled (4 °C) methanol:water (3:1, v/v) including 0.1 % formic acid (v/v) in an ultrasonic bath. After centrifugation an aliquot of the supernatants (300 µl of native samples and 420 µl of U-^{13}C labelled reference sample) were transferred separately to another 1.5 ml-Eppendorf tube and pre-cooled (4 °C) water + 0.1 % formic acid (v/v) was added to achieve a final methanol:water ratio of 1:1 (v/v). IS was achieved by adding the same volume of U-^{13}C labelled sample aliquots resulting in (1:1, v/v) mixtures of non-labelled and U-^{13}C labelled supernatant. All samples were rigorously mixed for 10 s before transfer into HPLC vials for LC–HRMS measurements.

Maize line CO354 ($n = 3$) and U-^{13}C labelled cultivar "Yukon chief" were prepared according to the same protocol, with "Yukon chief" diluted sample extracts being used for IS.

2.4 LC–HRMS analysis

All samples (*F. graminearum,* wheat, maize) were analysed on a UHPLC system (Accela, Thermo Fisher Scientific, San Jose, CA, USA) coupled to an LTQ Orbitrap XL (Thermo Fisher Scientific) equipped with an ESI source. A HTC PAL system (CTC analytics, Zwingen, Switzerland) was used for injection (10 µl) per sample and for thermostatisation of sample solutions to 10 °C throughout the whole sequence.

A reversed-phase XBridge C_{18}, 150 × 2.1 mm i.d., 3.5 µm particle size (Waters, Milford, MA, USA) analytical column, preceded by a C_{18} 4 × 3 mm i.d. security cartridge (Phenomenex, Torrance, CA, USA) was thermostated to 25 °C and used for chromatographic separation at a constant flow rate of 250 µl/min. Water containing 0.1 % FA (v/v) (eluent A) and MeOH containing 0.1 % FA (v/v) (eluent B) were used for linear gradient elution: The initial mobile phase composition (10 % eluent B) was held constant for 2 min, followed by a linear gradient to 100 % eluent B within 30 min. After a hold time of 5 min the column was re-equilibrated for 8 min at 10 % eluent B. A 10 µl sample loop was employed to maintain a constant injection volume.

The ESI interface was operated in positive ion mode with the following settings: sheath gas: 60 arbitrary units, auxillary gas: 15 arbitrary units, sweep gas: 5 arbitrary units, capillary voltage: 4 kV, capillary temperature: 300 °C. LTQ parameters were automatically tuned for maximum signal intensity of a 10 mg L^{-1} reserpine solution (Sigma Aldrich) as recommended by the instrument manufacturer. For measurements using the FT-Orbitrap in the fullscan mode, the automatic gain control was set to a target value of 3×10^5 and a maximum injection time of 500 ms was chosen. The mass spectrometer was operated in a scan range from *m/z* 100–1,000 with a resolving power setting of 60,000 FWHM (at *m/z* 400). Data were recorded using Xcalibur 2.1.0 (Thermo Fisher Scientific).

2.5 Data processing

For an efficient extraction of metabolite derived MS signals and analytical features, several consecutive data processing steps (illustrated in Fig. 1-3-a–f) were implemented as an expansion of the already released software version of MetExtract (Bueschl et al. 2012). The focus of this paper was laid on the detailed description of the complete workflow including the biological experiment, sample analysis and data processing as well as its application to samples of plants and fungi. The extended MetExtract 2.0 software, which is capable of performing all of the following data processing steps (2.5.1–2.5.7) will be published elsewhere together with the release of the programme.

2.5.1 *Pre-processing (Fig. 1-3-a)*

Measurement files were converted from acquired profile to centroid data and the mzXML format (Pedrioli et al. 2004) with the MSConvert programme from the freely available

ProteoWizard package v.3.0.3980 32-bit (Chambers et al. 2012).

2.5.2 MS signal pair filtering and clustering of m/z values (Fig. 1-3-b)

Each recorded MS scan was inspected for the typical SIL derived isotopic pattern as described previously in Bueschl et al. (2012). The intensity threshold of both the monoisotopic ^{12}C derived (M) and U-^{13}C derived (M′) MS signal was set to 5,000 counts in at least 3 scans. The maximum *m/z* deviation from postulated *m/z* values was set to 2.5 ppm and the isotopologue abundance error was set to ±20 %. Each MS signal pair, fulfilling the criteria was annotated with *m/z* of M, the charge number z, deduced from the SIL derived isotopic pattern, and the determined number of carbon atoms n_C, calculated from the *m/z* value difference between M and M′, and z. Extracted MS signal pairs were clustered together with hierarchical clustering to group redundantly extracted MS signal pairs of similar *m/z* value (i.e. MS signal pairs, which originate from the same chromatographic peak or structural isomers with identical sum formula). Hierarchical clustering was performed separately for all MS signals having the same number of carbon atoms and the same charge number. All clusters in the resulting tree, whose *m/z* values differed more than ±10 ppm were split into separate sub-clusters.

2.5.3 Feature pair picking (Fig. 1-3-c)

For each MS signal cluster, the algorithm of (Du et al. 2006) was utilised to inspect the XICs of both the monoisotopic ^{12}C and the corresponding U-^{13}C labelled ions for co-eluting and similarly shaped chromatographic peaks. For this purpose, a maximum retention time difference of ±15 scans was tolerated between chromatographic peaks in both XICs. Furthermore, the chromatographic peak profiles of the monoisotopic ^{12}C and the U-^{13}C labelled features were compared with the Pearson correlation coefficient and only those, with correlation coefficients greater than 0.5 were considered a valid feature pair derived from the SIL process. This data processing step resulted in a list of putative feature pairs (monoisotopic ^{12}C- and corresponding U-^{13}C labelled feature) with each feature pair being annotated with the *m/z* value of M, retention time (Rt), peak area, number of carbon atoms per ion (n_C), and charge state z.

2.5.4 De-isotoping (Fig. 1-3-d)

Compared to correctly paired features, M+1 features, falsely picked as monoisotopic ^{12}C features or M′−1 features, falsely picked as U-^{13}C labelled features showed a reduced number of carbon atoms n_C and/or an increased monoisotopic ^{12}C *m/z* value. Thus such erroneously extracted features were removed from the feature list by comparing the *m/z* values of M, charge state z, Rt and n_C among putative feature pairs.

2.5.5 Feature pair grouping (Fig. 1-3-e)

To group different features from the same metabolite, extracted feature pairs were convoluted by comparing the chromatographic peak shapes of all monoisotopic ^{12}C features eluting at approximately the same retention time (±10 scans) (Kuhl et al. 2012). A minimum correlation coefficient of 0.85 was specified for features to be grouped together.

2.5.6 Matching results of several samples and generation of data matrix (Fig. 1-3-f)

To track metabolite features over all samples of a particular experiment, the extracted feature pairs of all LC–HRMS data files were compared using n_C, *m/z* of M and Rt in that order. After data matrix generation, monoisotopic ^{12}C and U-^{13}C labelled features, initially missed in some of the data files due to the restrictive filtering criteria, were searched for in a targeted way. To this end, the described peak picking and integration algorithms were employed but without checking peak shape similarity.

2.5.7 Internal standardisation

Internal standardisation was carried out on a file basis for each feature pair by dividing the area of monoisotopic ^{12}C by that of its corresponding U-^{13}C labelled feature.

2.6 Comparison with labelling free strategy

To compare the feature extraction process with a labelling free approach, a non-labelled *F. graminearum* aggregate sample, which had been diluted with FMM (1:1, v/v) and did not contain any U-^{13}C labelled culture supernatant and one of the U-^{13}C standardised *F. graminearum* aggregate samples, were analysed, processed and evaluated. The data file derived from a non-labelled sample was processed with XCMS (1.34.0) and R (R Development Core Team, 2012, v. 2.15.2) using parameter settings as recommended by Patti et al. (2012a) for HPLC Orbitrap XL MS. The LC–HRMS data file obtained for the U-^{13}C standardised aggregate sample was processed as described above (steps 2.5.1–2.5.4) and parameter settings similar to XCMS (i.e. minimum intensity of 5,000 in at least 3 scans; maximum tolerated *m/z* deviation of 2.5 ppm). Automated comparison of the results was performed by comparing both the

determined *m/z* value and retention time of all extracted features and feature pairs respectively. For this, a maximum relative *m/z* deviation of ±10 ppm and ±0.15 min was allowed for two results to match. Features, which had only been found by the SIL assisted data processing, were further inspected manually using TOPPView (Sturm and Kohlbacher 2009, v 1.10).

2.7 Selectivity evaluation of SIL assisted workflow

To demonstrate the selectivity of feature pair extraction in the presented SIL assisted metabolomics workflow, blank samples (solvent blank) ($n = 3$) as well as five non-labelled *F. graminearum* aggregate samples (no internal standardisation with U-^{13}C labelled supernatant) were processed as described earlier.

2.8 Evaluation of internal standardisation and matrix effects

Analysis of internal standardisation was performed on a feature pair level using only those pairs for which both the monoisotopic ^{12}C and the corresponding U-^{13}C labelled features were found in all replicates of a certain sample type after re-integration. Therefore, no imputation of missing values was required. For analytical precision demonstration before and after internal standardisation coefficient of variance (CV) histograms of individual feature pairs within all replicates of a sample type were calculated. The bin width was set to 5 %. CV values above 120 % were set to 120 % to achieve equidistant axis in the plots. To demonstrate internal standardisation with multivariate statistics, PCA plots were calculated for the monoisotopic ^{12}C and U-^{13}C labelled feature areas as well as for the internal standardisation derived feature pairs. For analytical precision analyses R (R Development Core Team 2012 v. 2.15.2) was used. The functionality for calculating the principal component analysis (PCA) was taken from the package ChemometricsWithR (Wehrens 2011, pp. 53–57). Data were range scaled (van den Berg et al. 2006) prior to PCA. For the ellipsis in the PCA plots, the ellipse package (Murdoch and Chow 1996) was used. Ellipses were calculated using the co-variance matrices of PC1 and PC2 of the respective sample types.

3 Results and discussion

U-^{13}C or ^{15}N labelled metabolites show nearly identical physico-chemical properties as their native non-labelled analogues. As a consequence, LC–HRMS measurements of mixtures of non-labelled and U-^{13}C labelled biological samples result in perfect co-elution of all isotopologues of a particular metabolite with very similar chromatographic peak shapes (Fig. 2). Thus, the analysis of mixtures of native and U-^{13}C labelled biological samples leads to labelling-specific isotopic distributions of both the non-labelled and U-^{13}C labelled metabolites in all recorded mass spectra containing biologically derived ion signals. As can be expected from the ESI process, different ion species such as protonated molecules as well as sodium adducts or the loss of water from the intact molecules may be observed. For each of the detected ion species two distinct mirror-imaged isotopic patterns are present in the mass spectra. In addition to the regular signal pattern originating from the natural isotopic composition of carbon (98.8 % ^{12}C and 1.1 % ^{13}C), the second isotopic pattern shows ascending MS signal intensities towards higher *m/z* values for all U-^{13}C labelled metabolite derived ion species. The relative abundance of the isotopic signals in the pattern of the labelled metabolite is given by the degree of ^{13}C enrichment achieved in the respective experiment (variant A) or the labelled biological reference sample (variant B). For the cultivation of *F. graminearum* strains (variant A) the degree of ^{13}C enrichment of metabolites was estimated from one data file using highly abundant features of both the monoisotopic ^{12}C and the corresponding U-^{13}C labelled isotopologues. From the intensity ratio of $M'-1$ to M' as well as the deduced number of carbon atoms for this isotopic pattern, the enrichment was calculated to be as high as 99.5 %, which is in good agreement with the suppliers specifications (99 %). For the wheat and maize samples (variant B), the U-^{13}C labelled reference samples also corresponded well to the supplier's specifications of around 97.5 %.

Variant A can be realized without much extra effort for less complex organisms such as bacteria, yeasts or fungi, which can be grown on synthetic media that require only a limited number of carbon sources, available as stable isotope enriched nutrients. Growing U-^{13}C labelled plants (e.g. *A. thaliana*, wheat) has also been carried out successfully but is a more challenging task regarding infrastructure, costs, and time to establish cultivation conditions in a controlled U-$^{13}CO_2$ enriched atmosphere and can be demanding, particularly under the experimental conditions of interest. If U-^{13}C labelled plant material is commercially available, variant B might be a good compromise as it provides the possibility to use globally U-^{13}C labelled plant samples as reference for both qualitative and quantitative measurements with the drawback however, that the biological experiment itself is not carried out under labelling conditions. Mammalian organisms can be regarded to be even more demanding than plants. So far, successful SIL experiments (^{15}N or ^{13}C stable isotope enrichment >90 %) of different mammalian cell lines (e.g. CHO) requiring complex media have been reported (Egorova-Zachernyuk

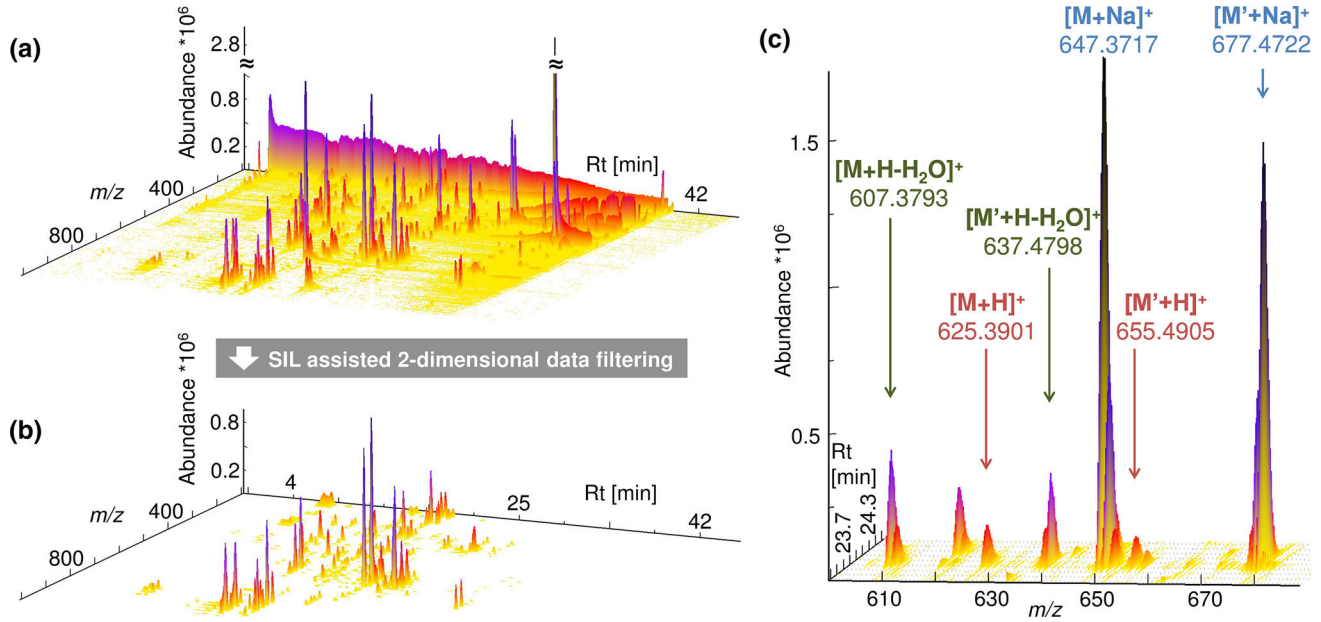


Fig. 2 3D representation of a selected *F. graminearum* aggregate sample analysed with LC–HRMS. Chromatogram of the unprocessed, centroided (**a**) and the processed (**b**) with only the SIL derived MS signals are shown. The 3D representation in c shows a zoomed section of the unprocessed datafile (**a**) illustrating the labelling specific isotopic pattern for three different ion species (*M* denotes the monoisotopic ^{12}C metabolite and *M′* denotes the U-^{13}C labelled metabolite) of a metabolite with the neutral, monoisotopic mass of 624.3827 u and $n_C = 30$ carbon atoms. 3D representations were created with TOPPView (Sturm and Kohlbacher 2009, v. 1.10) [figure-width: 174 mm]

et al. 2011), but to the best of our knowledge, labelling of whole mammalian organisms was not performed successfully to date.

With the presented approach, the systematic search of MS-signals and feature pairs carrying the SIL specific isotopic pattern virtually enables the complete annotation of that part of the metabolome of a biological sample which can be accessed by the chosen sample preparation and measurement method. Moreover, since the mass difference of the ^{12}C monoisotopic ion M and its U-^{13}C labelled isotopologue ion M′ is proportional to the mass difference between ^{12}C and ^{13}C isotopes (i.e. 1.00335 u), the number of carbon atoms contained in that particular metabolite ions can directly be calculated from the measured HRMS spectra (Bueschl et al. 2012). A prerequisite to unambiguously assign the number of carbon atoms (n_C) in a particular metabolite ion directly, the workflow requires the *m/z* value of both the monoisotopic, non-labelled M- and fully ^{13}C labelled M′ ions to be clearly identifiable among their respective isotopic patterns (Fig. 2c). For example, a ^{13}C enrichment of 98 % for the labelled metabolites allows extracting and directly calculating the number of carbon atoms in metabolite ions containing up to 60 carbon atoms. In case the degree of ^{13}C enrichment dropped to e.g. 85 %, this number is reduced to a maximum of 7 carbon atoms before the MS signal of a hypothetical $^{13}C\ ^{12}_{n-1}C_1$ reached the same intensity as the fully labelled $^{13}C_n$ isotopologue and thus would interfere with direct n_C assignment. It should be noted however that even at a further reduced enrichment degree of as low as e.g. 75 or 50 %, the presented workflow does still allow for the automated recognition of corresponding isotopic ion patterns when parameter settings for processing are adjusted accordingly. In addition to the qualitative aspects of improved metabolome annotation, the use of globally U-^{13}C labelled biological samples permits a highly efficient internal standardisation thereby enabling the assessment of precision parameters (both biological and technical) as well as compensation of matrix effects and improved relative quantification of hundreds of metabolites simultaneously.

3.1 Feature reduction by two-dimensional data filtering and feature grouping

With an average number of approximately 2 million signals (corresponding to 900 MS signals/mass scan), the raw chromatograms of *Fusarium* culture supernatants contained less data points than the plant derived chromatograms carrying roughly 3 million MS signals (1,200 MS signals/mass scan). This greater complexity of the plant samples is also visible in all successive data processing steps. The first data filtering step, comprises an inspection in every mass spectrum of a particular

Table 1 The table provides a quantitative overview of the data processing results with the proposed workflow (Fig. 1)

No.	Workflow step	Wildtype PH-1 (variant A)	Aggregate samples (variant A)	Remus wheat (variant B)	CM wheat (variant B)	CO354 maize (variant B)
MS signals	3-a	1.8×10^6 ($\pm 0.06 \times 10^6$)	1.9×10^6 ($\pm 0.09 \times 10^6$)	3.0×10^6 ($\pm 0.01 \times 10^6$)	3.0×10^6 ($\pm 2 \times 10^2$)	2.8×10^6 ($\pm 7 \times 10^3$)
SIL derived signal pairs	3-b	1.06×10^4 ($\pm 0.24 \times 10^4$)	1.65×10^4 ($\pm 0.52 \times 10^4$)	1.87×10^4 ($\pm 0.66 \times 10^4$)	2.09×10^4 ($\pm 0.19 \times 10^4$)	2.04×10^4 ($\pm 0.67 \times 10^4$)
MS signal clusters	3-b	8.09×10^2 ($\pm 0.27 \times 10^2$)	1.28×10^3 ($\pm 0.04 \times 10^3$)	2.07×10^3 ($\pm 0.03 \times 10^3$)	2.32×10^3 ($\pm 0.13 \times 10^3$)	1.39×10^3 ($\pm 0.37 \times 10^3$)
Feature pairs before de-isotoping	3-c	824 (±19)	1,199 (±34)	1,288 (±30)	1,443 (±92)	858 (±274)
De-isotoped feature pairs	3-d	291 (±9)	442 (±14)	797 (±9)	902 (±68)	511 (±166)
Feature groups (i.e. metabolites)	3-e	87 (±6)	135 (±6)	347 (±12)	362 (±32)	209 (±58)

Selected sample types are taken from the *F. graminearum*, wheat and maize experiment. The mean value and its standard deviation among the replicates within a certain sample type are given Table 1

data file for groups of corresponding M (i.e. monoisotopic ^{12}C metabolite ion), M+1, M′ (i.e. U-^{13}C labelled metabolite ion) and M‘−1 isotopologue MS signals forming MS signal pairs. The formation of MS signal pairs in each mass scan resulted in the most significant reduction of data points. As illustrated for a selected *F. graminearum* aggregate sample (respective numbers for other sample types and experiments can be found in Table 1), the LC–HRMS raw chromatogram (centroided and converted to mzXML) contained a total of 1,987,654 MS signals which were reduced by a factor of about 120 to 16,736 putative M/M′ signal pairs under the tested conditions (Fig. 2a, b). On average, the number of extracted signal pairs represents 0.6–0.9 % of the original contained MS signals. It should be noted that low abundant principal ions (i.e. M or M′) may not show distinctive M+1 or M′−1 isotopic signals (e.g. at the beginning/end of a chromatographic peak), and thus these signals are not considered during this filtering step. Therefore, the number of real SIL derived signal pairs is always underestimated. However, data processing which does not verify the isotopic patterns could successfully extract these low abundant metabolite ions, but at the same time increase the number of false positive findings (e.g. pairings of artefacts which can originate from the Fourier transformation process (Brown et al. 2009)). However, all SIL assisted data processing steps presented here always verify the isotopic patterns of both the native and U-^{13}C labelled metabolite ions using the number of carbon atoms for this ion species deduced from M and M′ respectively.

After MS signal clustering, feature pairs are extracted from the data. As the non-labelled monoisotopic and its corresponding U-^{13}C labelled analogue of a particular metabolite can be expected to show perfect chromatographic co-elution, verification of retention time and chromatographic peak shape similarity is used for feature pair picking (Fig. 1-3-c). The subsequent de-isotoping step eliminates incorrectly paired monoisotopic ^{12}C- and U-^{13}C features which do not represent true monoisotopic or uniformly labelled features. Together, the data filtering steps 3b–3d reduced the metabolite-related information—depending on the investigated organism—to ca. 300–900 distinct de-isotoped feature pairs per LC–HRMS chromatogram.

Since ionisation by electrospray may give rise to several ion species for the same substance such as adducts, in-source fragments or dimers, the SIL derived feature pairs are further combined with the aim to convolute all ion species of a particular metabolite into single groups (Fig. 1-3-e). Feature grouping is greatly facilitated by the prior removal of all non-biology related as well as all M+1, M+2 isotopic features and, depending on the investigated samples, resulted in the detection of 87–135 metabolites for the *F. graminearum*—and 200–360 truly plant derived substances for the maize and the wheat extracts respectively. As a major benefit of the presented approach all of these metabolites can be used to build-up reference databases and can serve as positive lists for future metabolomics experiments. The list of approximately 135 metabolites detected for the *Fusarium* aggregate samples (variant A, labelling under experimental conditions) can be assumed to comprise all metabolites which have been produced and released by at least one of the fungal strains under the tested conditions. For variant B of the workflow, which is exemplified with globally U-^{13}C-labelled reference samples, the detected metabolites are restricted to those present in both, the reference as well as the experimental samples. Nevertheless, all of the generated feature groups facilitate metabolic feature annotation and come with additional valuable characteristics for molecular formula generation and metabolite annotation such as the number of carbon atoms, the charge state for each metabolite ion as well as all other ion species detected for that particular metabolite (i.e. feature group).

3.1.1 *Comparison of feature extraction with a labelling-free workflow*

To compare the number of SIL derived features with those found with a labelling-free metabolomics approach, a selected native *F. graminearum* aggregate sample (consisting of a mixture of non-labelled *F. graminearum* wildtype and non-labelled *tri5*Δ mutant culture supernatants) was diluted 1:1 (v/v) with either *Fusarium* minimal medium (FMM) or a pooled ^{13}C-aggregate supernatant, to generate two LC–HRMS samples both containing the native metabolites at identical concentration levels. Retention time shifts observed between the respective LC–HRMS chromatograms were negligible (see TICs in Fig. 3a). Parameter settings for XCMS-based (Smith et al. 2006) feature extraction for the non-labelled aggregate sample chromatogram such as intensity threshold (5,000 counts), minimum number of scans (3×), maximum ppm deviation (2.5 ppm) were kept identical to the SIL assisted feature picking in order to ensure maximum comparability of results between the two approaches. As expected, with XCMS every chromatographic peak in the data regardless of its origin (biological, background, noise…) was extracted as a feature. In total, $n = 4{,}625$ features (illustrated as grey symbols in scatter plot of Fig. 3b) were found by XCMS based data processing including all low abundant features with no observable isotopic peaks or MS signals. In contrast, with the SIL assisted approach MS signal- and feature pair picking and subsequent de-isotoping only yields monoisotopic ^{12}C features with a high degree of confidence to correspond to truly sample derived metabolites. In total, application of the SIL assisted workflow yielded 431 feature pairs (red dots in scatter plot in Fig. 3b) which are about ten times less compared to XCMS. Moreover, data processing and automated comparison of the SIL and XCMS assisted approach resulted in 28 feature pairs (Fig. 3b blue dots) solely found with the SIL assisted approach. A closer, manual inspection of these 28 results showed, that 22 features were not automatically matched because of larger *m/z* or retention time deviations. Only three feature pairs found solely with the SIL assisted approach did not show any signals in the native sample

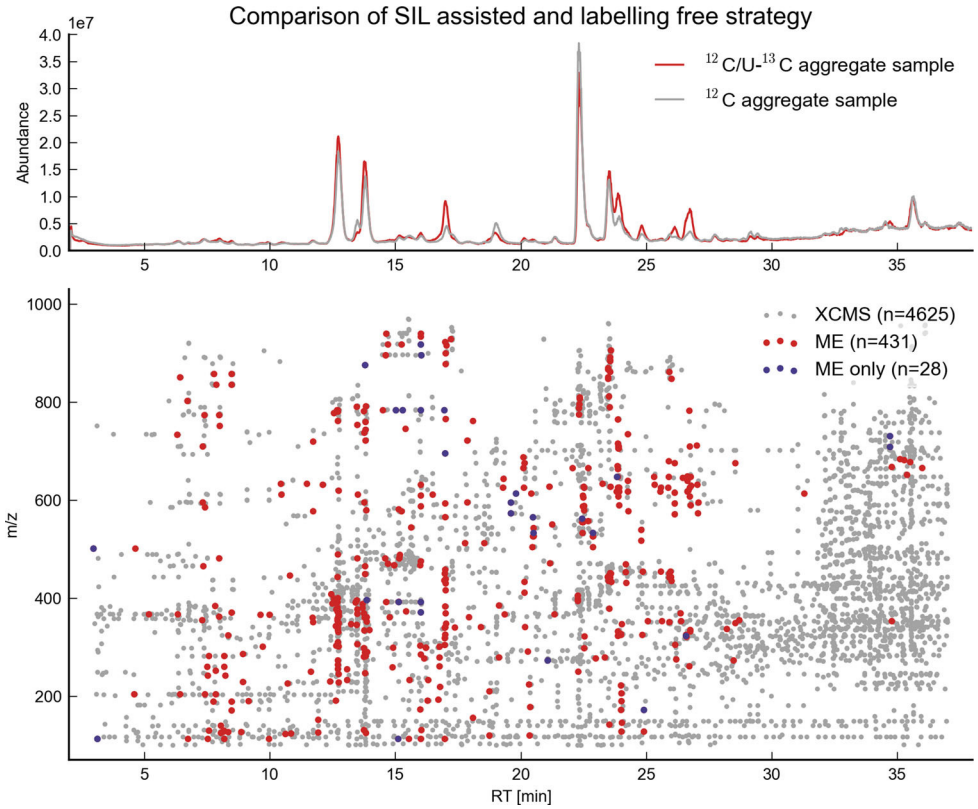


Fig. 3 **a** Illustration of an overlay of full scan LC–HRMS total ion current chromatograms obtained for two *F. graminearum* aggregate samples. *Red* Non-labelled ^{12}C and U-^{13}C culture filtrate mixed 1:1 (v/v); *grey* Non-labelled filtrate mixed 1:1 with fungal growth medium. **b** 2D plot of detected LC–HRMS features (*all dots*). *Grey symbols* indicate all features found with XCMS processing. *Red symbols* represent monoisotopic ^{12}C features found by both XCMS and the presented workflow (*variant A*, Fig. 1). Monoisotopic ^{12}C features found by the labelling assisted approach only are marked in *blue*. Features with a retention time >30 min are mainly detected by XCMS. Due to the higher strength of the eluent, predominantly impurities of non-biological origin such as polymers and apolar compounds are displaced from the stationary phase [figure-width: 174 mm]

processed with XCMS. Moreover, 3 of these 28 feature pairs were identified as false positives (e.g. pairings of Fourier transform artefacts (Brown et al. 2009)). For additional manual inspection of those parts of the LC–HRMS chromatograms showing a high density of features (e.g. Rt $\geq$ 30 min or *m/z* $\leq$ 150), which had only been found by XCMS, TOPPView was used to confirm that none of these features did show corresponding U-^{13}C labelled isotopologues.

It should be noted that workflows which do not make use of SIL, generally try to not further consider those non-biologically related background features by using statistical analysis to select features significantly differing between two or more experimental conditions. Such an assumption implies, however that only biologically derived features vary significantly between the different experimental conditions and that all others approximately show similar abundances across the different sample types.

3.1.2 Selectivity of the presented approach

To demonstrate the selectivity of the workflow for the extraction of truly biologically related metabolite features by the use of SIL, eight blank samples were included in a measurement sequence of *F. graminearum* samples and evaluated according to the presented workflow (variant A). Three of these blanks, which consisted of purified water instead of fungal supernatant, were employed as solvent blanks. Furthermore, five aggregate samples containing only the non-labelled *F. graminearum* metabolome served as simulated matrix background blanks. Metabolite ions detected in the native *F. graminearum* supernatants were not expected to be found with the SIL assisted data processing steps since they did not contain any U-^{13}C labelled metabolites. In the solvent blanks, hardly any MS signal pairs (searched in each mass scan, Fig. 1-3-b) were extracted (2, 26 and 39 MS signal pairs respectively). Better yet, none of these MS signal pairs were further confirmed to be valid feature pairs according to the predefined filtering criteria (Fig. 1-3-c). The simulated background blanks showed 14, 17, 7, 224 and 106 MS signal pairs on a mass scan level. Subsequent feature pair picking revealed 0, 1, 1, 1 and 5 feature pairs for the simulated background blanks respectively. However, none of such extracted feature pairs were found in more than one of the measured matrix blanks. Further manual inspection clearly showed that all of these randomly picked feature pairs fulfilled the present criteria either by chance or were pairings of different adducts or Fourier transformation artefacts (Brown et al. 2009). Such incorrectly picked adducts or artefacts showed nearly identical chromatographic profiles and they were therefore not discarded as false positives automatically (Fig. 1-3-e). Two of these feature pairs were detected at >700 u with a difference between the monoisotopic ^{12}C and corresponding U-^{13}C mass corresponding to less than ten carbon atoms. Thus such feature pairs can easily be excluded from further data analysis. In conclusion the very low rate of false positives in both types of blank samples demonstrates the exceptionally high selectivity of the presented approach in only extracting truly biologically related feature pairs.

3.2 Results of internal standardisation

Absolute quantification in untargeted metabolomics experiments of all detected (known or unknown) metabolites is generally not feasible by most of the current approaches. Moreover, the accuracy of relative feature/metabolite quantification in untargeted metabolomics experiments is limited by matrix effects which can cause problems during statistical analysis as the biased feature abundances complicate comparison across different experimental conditions. Internal standardisation using globally stable isotope labelled biological samples provides the ideal tool to overcome these limitations and has already been used for relative (e.g. Giavalisco et al. 2009) and even for absolute quantification in untargeted metabolomics approaches (Bennett et al. 2008). As the presented workflow also makes use of globally U-^{13}C labelled biological samples, the effect of global internal standardisation on matrix effects and technical precision has been investigated at the example of the *F. graminearum* dataset.

3.2.1 Correction of matrix effects

In many metabolomics experiments unsupervised multivariate statistical tools such as PCA are used as a first step to reduce the dimensionality of the analytical data and test for separation of biological samples into different classes according to experimental conditions. Such tools operate on different signal abundances or feature areas in the data matrix obtained from prior data processing. In order to test the suitability of the SIL assisted workflow to correct for matrix effects, feature areas obtained for the *F. graminearum* dataset were range scaled (van den Berg et al. 2006) and subsequently PCA score plots were calculated using (1) peak areas of monoisotopic ^{12}C features (^{12}C-PCA, Fig. 4a), (2) U-^{13}C labelled ion species (U-^{13}C-PCA, Fig. 4b) and (3) the peak area ratio of the respective non-labelled and U-^{13}C features of a certain feature pair (^{12}C/U-^{13}C-PCA, Fig. 4c). To be able to compare the three different PCA plots, only those feature pairs which had been found consistently throughout all replicates and samples categories (PH-1, *tri5*Δ and pooled aggregate samples (AGs)) were considered ($n = 109$).

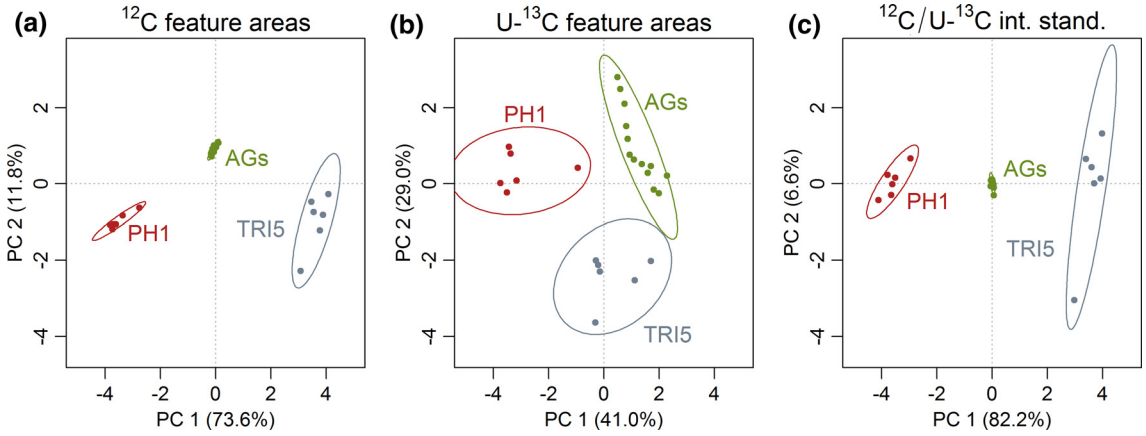


Fig. 4 Three PCA *scores plots* derived from consistently extracted feature pairs of three sample types: *F. graminearum* samples PH-1, *tri5*Δ and aggregate samples (AGs). For all three PCAs the exactly same set of feature pairs was used, however different intensity values (*peak areas*) were taken for each feature pair. **a** areas of monoisotopic ^{12}C features of the respective feature pairs, **b** areas of U-^{13}C labelled features, **c** intensity ratios of monoisotopic ^{12}C and corresponding U-^{13}C feature area (internal standardisation) [figure-width: 174 mm]

As can be expected from the experiment, the score plot of the ^{12}C-PCA shows a clear separation of the PH-1, *tri5*Δ and AGs samples into three distinct groups. The same holds true for the ^{12}C/U-^{13}C-PCA, which was calculated from the area ratios of non-labelled and corresponding U-^{13}C features. Additionally the variance captured by PC1 and PC2 increased slightly from 85.4 to 88.8 %. Compared to the ^{12}C-PCA, the aggregate samples are located in the centre of the ^{12}C/U-^{13}C-PCA score plot (Fig. 4c), which is explained by the range scaling process together with the fact that these aggregate samples constitute of equal amounts of U-^{13}C internally standardised PH-1 and *tri5*Δ samples. In contrast to the ideal behaviour of the U-^{13}C-standardised AG samples, matrix effects obviously affected the monoisotopic ^{12}C feature areas of the aggregate samples (Fig. 4a), which in turn become visible as a shift of the aggregate sample group away from the centre of the ^{12}C-PCA score plot.

Interestingly, the use of U-^{13}C-feature areas for PCA also resulted in a clear separation into the three sample categories, although (according to the preparation protocol for the AGs, see Sect. 2.3.1) the absolute concentration levels of all U-^{13}C labelled metabolites were identical in every of the analysed samples and sample type. In this case, the separation of the sample groups in the U-^{13}C-PCA plots is explained by the different metabolic composition of the wildtype PH-1, *tri5*Δ and AGs samples with respect to their non-labelled metabolites, which resulted in distinct alterations of the areas derived from U-^{13}C labelled features (i.e. matrix effects) for each of the tested sample categories. The separation of the three sample categories based on the peak areas of the respective U-^{13}C features would have never been recognised as artefact (caused by matrix effects) without the availability of globally U-^{13}C labelled biological samples. In contrast, an observation as depicted in Fig. 4b most probably would have led to the false conclusion that metabolites differing between the experimental samples had caused the separation according to the tested experimental states.

3.2.2 *Precision of workflow and improvement of technical data variability*

Stable isotope labelling assisted internal standardisation has been successfully used for improved metabolite quantification in GC–MS and LC–MS based metabolomics studies (e.g. (Birkemeyer et al. 2005; Bennett et al. 2008; Giavalisco et al. 2009)). Here, the assessment and improvement of both biological and technical precision of the presented SIL assisted workflow are exemplified with the *F. graminearum* dataset. Again, U-^{13}C labelled fungal samples were employed for global internal standardisation of non-labelled samples and subsequently precision measures of the workflow were estimated. Only those feature pairs were considered which had been found in all replicates of the respective sample type corresponding to $n = 307$ ^{12}C/U-^{13}C feature pairs for PH-1 wildtype data and $n = 424$ for aggregate samples. Coefficients of variation (CVs) of each feature (pair) were calculated across all replicates of a particular sample type, the distribution of CVs was plotted as a histogram with a class size of 5 % (Fig. 5) and the median CV as well as 90 % percentile were taken as precision estimate. For PH-1 samples, CVs of monoisotopic ^{12}C feature areas showed a median value of 15.1 % with 90 % of all features showing CVs below 36 %. These CV values can be interpreted as a measure for the variability of the overall workflow including all steps from culturing of fungi (biological variance) to sample

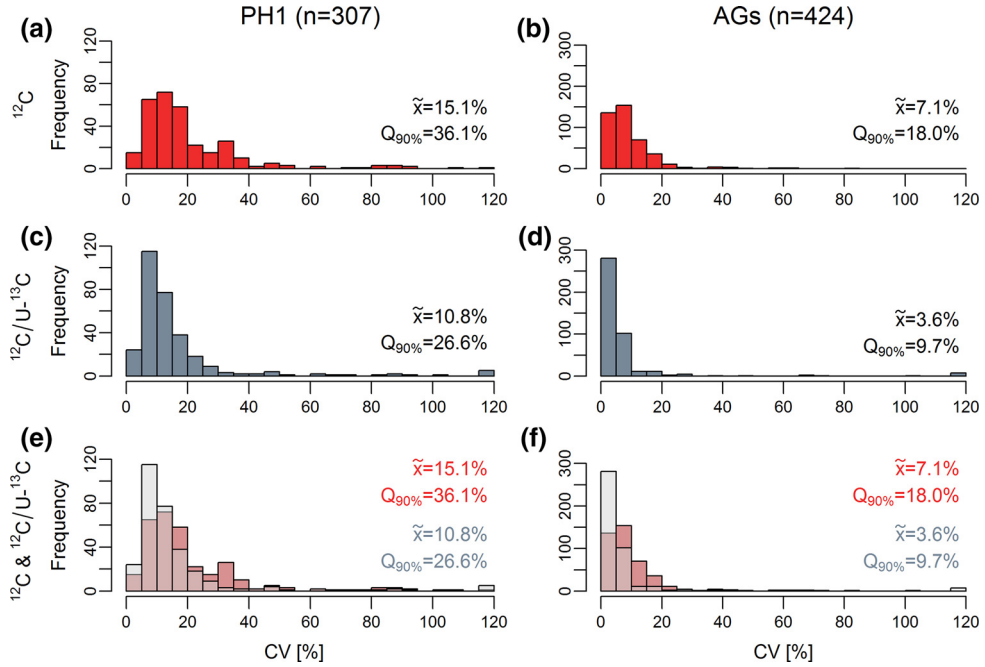


Fig. 5 Histograms showing the distributions of coefficients of variation (CV) across all SIL derived features which were consistently found in all replicates of *F. graminearum* wildtype PH-1 ($n = 6$) and *F. graminearum* aggregate samples ($n = 13$). The histograms in **a** and **b** (*red*) were derived from the peak areas of the monoisotopic ^{12}C feature of the respective feature pairs while **c** and **d** (*blue*) were calculated after internal standardisation with the areas of the corresponding U-^{13}C labelled features of the very same feature pair. Histograms **e** and **f** (overlay of transparent *red* and *blue*) combine the respective above two histograms to illustrate the shift towards lower CVs by internal standardisation, achieved for both sample types [figure-width: 129 mm]

preparation, LC–HRMS measurement, data processing and feature integration (technical variance). Using the area ratio of corresponding monoisotopic ^{12}C and U-^{13}C features, the overall precision of PH-1 samples was improved to a median CV of 10.8 %, with 90 % of all features showing CVs below 26.6 %, indicating that (1) biological and technical variability roughly contributed equally to the overall spread of feature areas and (2) internal standardisation with globally U-^{13}C labelled samples helped to improve precision considerably. For the *tri5*Δ mutant, comparable CV values and improvements were obtained (data not shown).

In order to estimate and dissect the precision of the end determination step (i.e. LC–HRMS measurement, data processing and feature integration), ^{12}C/U-^{13}C aggregate samples, consisting of 1:1 mixtures of PH-1/*tri5*Δ supernatants were measured as replicate injections ($n = 13$) at regular intervals over a complete LC–HRMS sequence. The distribution of CVs of the monoisotopic ^{12}C feature areas showed a median CV of 7.1 % (90 % of all features had CV values below 18 %). After internal standardisation (blue histogram), the distribution of CVs shifted left (blue histogram) towards lower CV values with a median and 90 % quartile of to 3.6 and 9.7 % respectively, demonstrating a substantial improvement (∼50 %) of the LC–HRMS end determination.

It should be noted that for a few ($n < 10$) features the internal standardisation of *F. graminearum* culture supernatants resulted in CV values >120 %. This might have been caused by non-reproducible degradation/chemical conversion of a few metabolite features in the samples but this phenomenon was not further investigated in this study however.

With ca. 20 % and 40–50 % respectively the distribution of CVs in wheat and maize samples (variant B) yielded slightly higher median and 90 % percentile values than for the *F. graminearum* samples (data not shown). Similar to the *F. graminearum* experiment, wheat and maize aggregate samples were used to study the technical precision. Similar to *Fusarium*, median CV (90 % percentile) values shifted from roughly 8 % (16 %) to 5 % (20 %) for wheat and 12 % (22 %) to 6 % (17 %) for maize respectively (data not shown).

In conclusion, the above illustrated results are in good agreement with the reports of e.g. Birkemeyer et al. (2005), Bennett et al. (2008), Giavalisco et al. (2009), who described enhanced precision and (relative) quantification after metabolome wide internal standardisation by use of U-^{13}C labelled biological samples. Furthermore, as shown for *F. graminearum* samples, internal standardisation resulted in an improved performance of multivariate data analysis.

4 Concluding remarks

Recent reports of SIL assisted tools and techniques and their application to various fields of metabolomics have illustrated significant improvements to circumvent major challenges of untargeted metabolite profiling. The presented workflow enables the untargeted global extraction of truly metabolite related MS signals and features in LC–HRMS datafiles derived from native and U-^{13}C labelled metabolomes. Together with the automated generation of hundreds of feature groups per sample, each of which is representing a distinct metabolite, this approach constitutes a major step forward towards global annotation of the entire metabolic composition of biological samples. Additionally, metabolome wide internal standardisation with U-^{13}C labelled samples greatly enhances accuracy and reliability of relative quantification by correction of technical variability as well as correction of matrix effects, which otherwise are difficult to evaluate and compensate.

In conclusion, although stable isotope labelling of whole metabolomes for untargeted metabolomics is still challenging and generally requires additional efforts in terms of costs and/or experimental design, it is anticipated that SIL assisted metabolomics will arouse increasing interest and become a well-established technique in metabolomics research.

Acknowledgments The authors thank Nora Neumann, Denise Schöfbeck, Sylvia Lehner and Benedikt Warth for useful discussions. The Austrian Science Fund (SFB-Fusarium-3706-B11;3702), WWTF (Toxi-Genome-9793008037), EC (KBBE-2007-22269-2-MYCORED), FFG (Grant-824186), BMWFJ, BMVIT, ZIT, and Zukunftsstiftung Tirol are acknowledged for financial support. V.M. was supported by Doctoral-School on the Agro-Food System (Agrisystem) of Università Cattolica del Sacro Cuore (Italy).

References

Bennett, B. D., Yuan, J., Kimball, E. H., & Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols, 3*, 1299–1311.

Birkemeyer, C., Luedemann, A., Wagner, C., Erban, A., & Kopka, J. (2005). Metabolome analysis: The potential of in vivo labeling with stable isotopes for metabolite profiling. *Trends in Biotechnology, 23*, 28–33.

Böttcher, C., Roepenack-Lahaye, E. V., Willscher, E., Scheel, D., & Clemens, S. (2007). Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry. *Analytical Chemistry, 79*, 1507–1513.

Brown, M., Dunn, W. B., Dobson, P., Patel, Y., Winder, C. L., Francis-Mcintyre, S., et al. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst, 134*, 1322–1332.

Bueschl, C., Kluger, B., Berthiller, F., Lirk, G., Winkler, S., Krska, R., et al. (2012). MetExtract: A new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. *Bioinformatics, 28*, 736–738.

Cano, P. M., Jamin, E. L., Tadrist, S., Bourdaud'Hui, P., Pean, M., Debrauwer, L., et al. (2013). New untargeted metabolic profiling combining mass spectrometry and isotopic labeling: Application on *Aspergillus fumigatus* grown on wheat. *Analytical Chemistry, 85*, 8412–8420.

Castillo, S., Gopalacharyulu, P., Yetukuri, L., & Orešič, M. (2011). Algorithms and tools for the preprocessing of LC–MS metabolomics data. *Chemometrics and Intelligent Laboratory Systems, 108*, 23–32.

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology, 30*, 918–920.

Chokkathukalam, A., Jankevics, A., Creek, D. J., Achcar, F., Barrett, M. P., & Breitling, R. (2013). mzMatch–ISO: An R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics, 29*, 281–283.

Clasquin, M. F., Melamud, E., & Rabinowitz, J. D. (2012). LC-MS data processing with MAVEN: A metabolomic analysis and visualization engine. *Current Protocols in Bioinformatics, 37*, 14.11.1–14.11.23.

Covey, T. R., Thomson, B. A., & Schneider, B. B. (2009). Atmospheric pressure ion sources. *Mass Spectrometry Reviews, 28*, 870–897.

de Jong, F. A., & Beecher, C. (2012). Addressing the current bottlenecks of metabolomics: Isotopic ratio outlier analysis^{TM}, an isotopic-labeling technique for accurate biochemical profiling. *Bioanalysis, 4*, 2303–2314.

de Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols, 2*, 778–791.

Du, P., Kibbe, W. A., & Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics, 22*, 2059–2065.

Dunn, W., Erban, A., Weber, R. M., Creek, D., Brown, M., Breitling, R., et al. (2013). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics, 9*, 44–66.

Egorova-Zachernyuk, T. A., Bosman, G. J., & Degrip, W. J. (2011). Uniform stable-isotope labeling in mammalian cells: Formulation of a cost-effective culture medium. *Applied Microbiology and Biotechnology, 89*, 397–406.

Giavalisco, P., Hummel, J., Lisec, J., Inostroza, A. C., Catchpole, G., & Willmitzer, L. (2008). High-resolution direct infusion-based mass spectrometry in combination with whole 13C metabolome isotope labeling allows unambiguous assignment of chemical sum formulas. *Analytical Chemistry, 80*, 9417–9425.

Giavalisco, P., Köhl, K., Hummel, J., Seiwert, B., & Willmitzer, L. (2009). 13C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography–mass spectrometry-based metabolomic research. *Analytical Chemistry, 81*, 6546–6551.

Hegeman, A. D. (2010). Plant metabolomics: Meeting the analytical challenges of comprehensive metabolite analysis. *Briefins in Functional Genomics, 9*, 139–148.

Hegeman, A. D., Schulte, C. F., Cui, Q., Lewis, I. A., Huttlin, E. L., Eghbalnia, H., et al. (2007). Stable isotope assisted assignment of elemental compositions for metabolomics. *Analytical Chemistry, 79*, 6912–6921.

Hiller, K., Metallo, C. M., Kelleher, J. K., & Stephanopoulos, G. (2010). Nontargeted elucidation of metabolic pathways using

stable-isotope tracers and mass spectrometry. *Analytical Chemistry, 82*, 6621–6628.

Hiller, K., Metallo, C., & Stephanopoulos, G. (2011). Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics. *Current Pharmaceutical Biotechnology, 12*, 1075–1086.

Hiller, K., Wegner, A., Weindl, D., Cordes, T., Metallo, C. M., Kelleher, J. K., et al. (2013). NTFD-a stand-alone application for the non-targeted detection of stable isotope-labeled compounds in GC/MS data. *Bioinformatics, 29*, 1226–1228.

Keller, B. O., Sui, J., Young, A. B., & Whittal, R. M. (2008). Interferences and contaminants encountered in modern mass spectrometry. *Analytica Chimica Acta, 627*, 71–81.

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., & Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *Journal of the American Society for Mass Spectrometry, 11*, 942–950.

Klein, S., & Heinzle, E. (2012). Isotope labeling experiments in metabolomics and fluxomics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 4*, 261–272.

Kluger, B., Bueschl, C., Lemmens, M., Berthiller, F., Häubl, G., Jaunecker, G., et al. (2012). Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat. *Analytical and Bioanalytical Chemistry, 405*, 5031–5036.

Kluger, B., Zeilinger, S., Wiesenberger, G., Schöfbeck, D., & Schuhmacher, R. (2013). Detection and identification of fungal microbial volatile organic compounds by HS–SPME–GC–MS. In V. K. Gupta, M. G. Tuohy, M. Ayyachamy, K. M. Turner, & A. O'donovan (Eds.), *Laboratory Protocols in Fungal Biology*. New York: Springer.

Koal, T., & Deigner, H. P. (2010). Challenges in mass spectrometry based targeted metabolomics. *Current Molecular Medicine, 10*, 216–226.

Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry, 84*, 283–289.

Leslie, J. F., & Summerell, B. A. (2007). *Media—Recipes and preparation. The Fusarium Laboratory Manual*. Hoboken: Blackwell Publishing.

Lommen, A., & Kools, H. J. (2012). MetAlign 3.0: Performance enhancement by efficient use of advances in computer hardware. *Metabolomics, 8*, 719–726.

Murdoch, D. J., & Chow, E. D. (1996). A graphical display of large correlation matrices. *American Statistician, 50*, 178–180.

Nakabayashi, R., Sawada, Y., Yamada, Y., Suzuki, M., Hirai, M. Y., Sakurai, T., et al. (2013). Combination of liquid chromatography–fourier transform ion cyclotron resonance-mass spectrometry with ^{13}C-labeling for chemical assignment of sulfur-containing metabolites in onion bulbs. *Analytical Chemistry, 85*, 1310–1315.

Patti, G. J., Tautenhahn, R., & Siuzdak, G. (2012a). Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nature Protocols, 7*, 508–516.

Patti, G. J., Yanes, O., & Siuzdak, G. (2012b). Innovation: Metabolomics: The apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology, 13*, 263–269.

Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., et al. (2004). A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology, 22*, 1459–1466.

Pluskal, T., Castillo, S., Villar-Briones, A., & Oresic, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics, 11*, 395.

R Development Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. Retrieved from http://www.R-project.org.

Redestig, H., Kobayashi, M., Saito, K., & Kusano, M. (2011). Exploring matrix effects and quantification performance in metabolomics experiments using artificial biological gradients. *Analytical Chemistry, 83*, 5645–5651.

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., van Ommen, B., et al. (2009). Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics, 5*, 435–458.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry, 78*, 779–787.

Sturm, M., & Kohlbacher, O. (2009). TOPPView: An open-source viewer for mass spectrometry data. *Journal of Proteome Research, 8*, 3760–3763.

Tang, L., & Kebarle, P. (1993). Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. *Analytical Chemistry, 65*, 3654–3668.

Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics, 9*, 504.

Theodoridis, G. A., Gika, H. G., Want, E. J., & Wilson, I. D. (2012). Liquid chromatography–mass spectrometry based global metabolite profiling: A review. *Analytica Chimica Acta, 711*, 7–16.

Trotzmüller, M., Guo, X., Fauland, A., Köfeler, H., & Lankmayr, E. (2011). Characteristics and origins of common chemical noise ions in negative ESI LC–MS. *Journal of Mass Spectrometry, 46*, 553–560.

Tulipani, S., Llorach, R., Urpi-Sarda, M., & Andres-Lacueva, C. (2013). Comparative analysis of sample preparation methods to handle the complexity of the blood fluid metabolome: When less is more. *Analytical Chemistry, 85*, 341–348.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

Vogeser, M., & Seger, C. (2010). Pitfalls associated with the use of liquid chromatography–tandem mass spectrometry in the clinical laboratory. *Clinical Chemistry, 56*, 1234–1244.

Wehrens, R. (2011). *Chemometrics with R: Multivariate data analysis in the natural sciences and life sciences*. Heidelberg: Springer.

Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst, 137*, 293–300.

Zhou, B., Xiao, J. F., Tuli, L., & Ressom, H. W. (2012). LC–MS-based metabolomics. *Molecular BioSystems, 8*, 470–481.

Publication #4: Tracing flavonoid degradation in grapes by MS filtering with stable isotopes

Alexander W. Chassy, **Christoph Bueschl**, Hyeyoung Lee, Larry Lerno,
Anita Oberholster, Daniela Barile, Rainer Schuhmacher, Andrew L. Waterhouse
Food Chemistry. 2015;166:448-455

Contributions of the presenting author: The author of this thesis performed data evaluation with the updated MetExtract software and performed statistical investigation for putative flavonoid degradation. Additionally, the author of this thesis wrote those parts of the manuscript detailing the data processing with MetExtract and statistical evaluation.

Supplementary material: Certain parts of the manuscript were transferred to supplementary material S1, which is not included in this work but can be accessed online at the journal's homepage (https://dx.doi.org/10.1016/j.foodchem.2014.06.002).

Food Chemistry 166 (2015) 448–455

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Tracing flavonoid degradation in grapes by MS filtering with stable isotopes

Alexander W. Chassy [a,1], Christoph Bueschl [b], Hyeyoung Lee [c], Larry Lerno [a], Anita Oberholster [a], Daniela Barile [c], Rainer Schuhmacher [b], Andrew L. Waterhouse [a,*]

[a] Department of Viticulture and Enology, University of California Davis, Davis, CA 95616, USA
[b] Department for Agrobiotechnology (IFA-Tulln), Center for Analytical Chemistry, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Str. 20, 3430 Tulln, Austria
[c] Department of Food Science and Technology, University of California Davis, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 10 March 2014
Received in revised form 23 May 2014
Accepted 1 June 2014
Available online 9 June 2014

Chemical compounds studied in this article:
Resveratrol (PubChem CID: 445154)
Malvidin-3-glucoside (PubChem CID: 443652)
Malvidin 3-(6″-acetylglucoside) (PubChem CID: 72193646)
Malvidin 3-(6″-p-coumarylglucoside) (PubChem CID: 72193651)
Quercetin 3′-O-glucoside (PubChem CID: 54758678)
Scirpusin A (PubChem CID: 5458896)
Petunidin 3-glucoside (PubChem CID: 443651)

Keywords:
Anthocyanin degradation
Metabolomics
Isotope tracer
Phenylpropanoid

ABSTRACT

Anthocyanin degradation has been proposed as one of the primary causes for reduced colour and quality in red wine grapes grown in a warm climate. To study anthocyanin degradation we infused berries with L-phenyl-$^{13}C_6$-alanine and then tracked the fate of the anthocyanins comparing normal (25 °C) and warm (45 °C) temperature conditions. An untargeted metabolomics approach was aided by filtering the MS data using software algorithms to extract all M and M+6 isotopic peak pairs, allowing the analysis to focus solely on the metabolites of phenylalanine. A paired-comparison *t*-test was performed over the 8 biological replicates revealing 13 metabolites that were statistically different between 25 °C and 45 °C treatments. Most of these features had lower abundances in 45 °C samples, confirming that 45 °C treatment caused anthocyanin degradation. In addition, resveratrol was significantly reduced following heat treatment. However, 5 metabolites increased following the 45 °C treatment. These unidentified metabolites are therefore suspects for anthocyanin degradation products.

1. Introduction

Anthocyanins are flavonoids responsible for blue to red pigments in strawberries, blueberries, grapes, and many other fruits and vegetables. In addition, anthocyanin metabolites are thought to be beneficial in the human diet (Forester & Waterhouse, 2010). In wine grapes, anthocyanin content and colour is considered highly reflective of quality (Somers, 1998). For example, the central valley of California produces more red wine grapes than any other part of the country but the price for these grapes is significantly less than the price for grapes produced in cooler wine regions. One of the primary reasons for this disparity is reduced anthocyanin content. Researchers have long known that warm climate causes lower anthocyanin content in central valley grapes, however, the mechanism was poorly understood (Winkler, Cook, Kliewer, & Lider, 1962). On the other hand, cool climates may also result in lower anthocyanin content (Ferrer-Gallego, Hernández-Hierro, Rivas-Gonzalo, & Escribano-Bailón, 2012). It appears that there may be an optimal temperature range for anthocyanin accumulation, but further conditions such as genotype, rootstock, soil, as well as others are also critical in modulating this optimum.

* Corresponding author. Tel.: +1 530 7524777; fax: +1 530 7520382.
E-mail address: alwaterhouse@ucdavis.edu (A.L. Waterhouse).
[1] Present address: National Institute of Health West Coast Metabolomics Center, University of California Davis, Davis, CA 95616, USA.

http://dx.doi.org/10.1016/j.foodchem.2014.06.002
0308-8146/

Recently, anthocyanin catabolism as a result of warm climate has been identified as an important influence of total anthocyanin content, yet to what extent remains unknown (Mori, Goto-Yamamoto, Kitayama, & Hashizume, 2007). Using freshly harvested grapes exposed to ^{13}C labelled phenylalanine *in vitro*, researchers observed increased degradation of labelled anthocyanins at elevated temperatures. Similarly, anthocyanin degradation has been observed in red grape cell suspension cultures, where anthocyanin loss was stemmed by treatment with magnesium (Sinilal et al., 2011). While the molecular mechanism of this degradation is still unknown, it has been demonstrated *in vivo*, that anthocyanin turnover is a crucial parameter in this respect. In principal, anthocyanin turnover may occur from an enzymatic pathway or non-enzymatic chemical modification, and a few studies have sought to uncover this phenomenon (Mori et al., 2007; Rowan et al., 2009). With respect to enzymatic metabolism, two potential pathways may result in the depletion of anthocyanin pools: an oxidative pathway and a deconstructive pathway. An oxidative pathway would involve enzymes such as polyphenol oxidase or peroxidases whereas a deconstructive pathway would simply involve movement of anthocyanins backwards through their biosynthetic pathways, perhaps starting with β-glucosidases (Sinilal et al., 2011). Evidence for both metabolic routes exist (Mori et al., 2007; Rowan et al., 2009). To our knowledge there is no published investigation of grape anthocyanin degradation *in vivo*. In order to improve anthocyanin content in warm climate grapes, a fundamental understanding of the turnover/degradation of anthocyanins must be established.

Phenylalanine has long been used to elucidate biochemical pathways (Neish, 1960), and in recent years mass spectrometry based studies greatly contributed to a better understanding of the kinetics and direction of phenylpropanoid metabolism (Boatright et al., 2004; Matsuda, Morino, Miyashita, & Miyagawa, 2003). In the last ten years mass spectrometry based metabolomics studies have begun to aim at the global detection of the entire catalog of natural products of a biological system. This latest of the -omics disciplines shows high potential for the global study of specific changes in metabolites. However with respect to anthocyanin degradation, only a few studies have been published so far. Bar-Akiva and colleagues, for example, applied an untargeted metabolomics approach and found 17 molecular features that differed between Brunsfelsia flower petals naturally transforming from purple to white (Bar-Akiva et al., 2010). They concluded that future studies might largely benefit from the use of isotopic tracers to help determine the formation and biochemical fate of labelled anthocyanins.

The combination of *in vivo* stable isotopic labelling with liquid chromatography-high resolution mass spectrometry (LC-HRMS) based untargeted metabolomics approaches offers great potential for the study of plant metabolism including phenylpropanoid and anthocyanin related metabolic pathways. Comprehensive labelling with $^{13}CO_2$ has been successfully used in combination with LC-HRMS for example to annotate the metabolome of *Arabidopsis thaliana* (Giavalisco et al., 2011; Hegeman et al., 2007) or to elucidate the diversity of anthocyanins in red cabbage (Charron et al., 2008). Apart from experiments aiming at the global labelling of whole plants, stable isotope tracer-assisted techniques have been shown to be powerful tools for the investigation of secondary metabolic pathways including phenylpropanoids by metabolic flux analysis (Heinzle, Matsuda, Miyagawa, Wakasa, & Nishioka, 2007; Matsuda et al., 2003). Recent improvements of analytical instrumentation such as HRMS and software algorithms for automated data processing enable the sensitive detection of metabolism products of xenobiotics and other secondary metabolites by untargeted metabolomics approaches. In a recent study, Kluger et al. described the automated and untargeted LC-HRMS based detection of novel mycotoxin conjugates formed in flowering wheat ears upon treatment with a mixture of the U-^{13}C labelled and native tracer toxin deoxynivalenol (Kluger et al., 2012). That study made use of the automated search for labelling specific isotopic patterns in the mass spectra by a novel data processing algorithm (Bueschl et al., 2012). A similar approach shall be used in the present work, looking at the constitutive phenylalanine metabolic pathways.

2. Materials and methods

2.1. Sample collection and preparation

A single cluster of self-rooted Cabernet Sauvignon, Clone 8, berries was removed from the Hopkins experimental vineyard at the University of California, Davis on August 19, 2011. At this time, approximately 75% of the berries in the vineyard had begun to ripen, turning from green to red. The grapes were then individually removed from the cluster by cutting the rakis. Once removed, the berries were sorted into 8 similarly sized and coloured groups of 4 berries. The berries of each group were then washed with sodium hypochlorate and ethanol, as previously described, and cut in half with a scalpel (Mori et al., 2007). Each half of each berry was separated into a treatment and control Petri-dish lined with filter paper, forming 8 replicate pairs of 4 berry halves each on a treatment and control Petri-dish for a total of 32 divided berries. Grape seeds were left intact in the grape unless they interfered with the splitting of berry halves, in which case they were removed. The Petri-dishes and filter paper were immediately hydrated with 2 mL of a filtered solution of 2 mM L-phenyl-$^{13}C_6$-alanine (Cambridge Isotope Labs, Cambridge, MA), 0.3 M sucrose in water. Each plate was held at ambient temperature under fluorescent light with 0.8 mL of $^{13}C_6$-Phe/sucrose added after 24 h to keep the plates hydrated. At 48 h, the grapes were transferred to new Petri-dishes with new filter paper, to which 2 mL of 0.3 M sucrose was added immediately and 0.8 and 0.125 mL were added at 72 and 96 h, respectively. Then after 96 h, the treatment and control groups were split into dark chambers held at 45 °C and 25 °C, respectively. Finally at 120 h, the grapes were frozen in liquid nitrogen and held at -80 °C until extraction.

Grape berry skins were separated from the pulp as they thawed and added to 10 mL MeOH (4 berry skin halves/tube). That mixture was sonicated in an ice bath for 2 h, after which an aliquot was diluted 50:50 (v/v) with water and successively filtered with 0.22 μm PTFE (Agilent, Santa Clara, CA) and 3000 kDa Amicon Ultra regenerated cellulose filters (Millipore, Cork, IR). The filtrate was then diluted 20:80 (v/v) with mobile phase A (see Section 2.2) and analysed by LC–MS and LC–MS/MS during separate injections.

2.2. ChipLC–MS and ChipLC–MS/MS

An Agilent 6520 Q-TOF MS coupled with an Agilent 1200 series HPLC-chip system (Agilent Technologies, Inc., Santa Clara, CA) was used in this study. A nano-HPLC chip (G4240-65010 UHC small molecule chip, Agilent Technologies, Inc., Santa Clara, CA) with a 500 nL enrichment column and 150 mm × 0.075 mm i.d. analytical column was used for chromatographic separation. Both the enrichment and analytical columns were packed with Zorbax 80SB-C18 (5 μm particle) stationary phase. The binary mobile phase consisted of (A) 3% acetonitrile/water (v/v) and (B) 90% acetonitrile/water (v/v), with both containing 0.1% formic acid. For loading of the sample onto the enrichment column, the capillary pump was operated at 4 μL/min using 99% solvent A. The sample on the enrichment column was transferred to the analytical column in a forward flush mode. A gradient-based chromatographic separation was performed on the analytical column and was driven

by the nanoliter pump running at 400 nL/min. The 27 min gradient used for the separation is as follows: 4–10 min, 1–15% B; 10–16 min, 15–25% B; 16–21 min, 25–80% B; 21–24 min, 80% B; 24–27 min, 1% B and a 5 min post equilibration time at 1% B.

The Q-TOF MS was operated in positive electrospray ionisation (ESI) mode for MS scan and data-dependent MS/MS. The source and ion optic parameters were optimised for minimal in-source fragmentation and improved sensitivity. The drying gas temperature and gas flow were 325 °C and 4 L/min, respectively. Capillary voltage was 1975 V for the entire LC run, and fragmentor voltage was set to 135 V. To cover the majority of *Vitis vinifera* flavonoids but exclude higher mass proanthocyanins which could contribute to background noise, recorded mass ranges were from *m/z* 105–700 for MS only and *m/z* 50–700 for MS/MS. Acquisition rates were 1 spectrum/s for MS and 3 spectra/s for MS/MS. The instrument was calibrated with ESI tuning mix (G1969-85000, Agilent Technologies, Inc., Santa Clara, CA), with reference masses at *m/z* 622.029 for positive mode. For the data-dependent MS/MS analysis, the precursor ion was automatically selected based on abundances with doubly charged ion being given the first priority followed by singly charged ion. The isolation width of automatically selected ions was set to ±2 u. The collision energy applied was based on the *m/z* of the ions and varied according to the following equation:

$$\text{Collision energy} = \frac{m/z}{100} \times 3.94 - 0.72$$

The optimum conditions for collision-induced dissociation (CID) were investigated through preliminary experiments.

2.3. MetExtract data analysis

2.3.1. Screening of labelling-derived MS signals

Each LC-HRMS raw chromatogram was processed with an extended but yet unpublished version of MetExtract (Bueschl et al., 2012) with the aim to automatically extract all M and M+6 peak feature pairs (see section results and discussion) containing the $^{13}C_6$ phenyl moiety. First each mass spectrum was inspected for candidate ^{12}C monoisotopic ions (M) and the associated characteristic isotopic pattern expected for metabolite ions, having the $^{13}C_6$ aromatic ring incorporated (M+6). Parameters for MS signal extraction consisted of a maximum tolerated mass deviation of ±20 ppm from the mass of the $^{13}C_6$ labelled ion (*m/z* value of M+6 was predicted from measured *m/z* of a putative M). Additionally, the presence of each of the M+1, M+5 and M+7 ions and their relative isotopic abundance ratios (deduced from highly abundant ^{12}C and corresponding $^{13}C_6$ isotopologues) were verified and used as a qualifier criterion for M and M+6 respectively. Finally, all M and M+6 MS signal pairs were grouped according to *m/z* values of M using hierarchical clustering with a maximum relative mass difference between the highest and lowest *m/z* value within the same MS signal cluster of ± 40 ppm.

2.3.2. Chromatographic peak picking, extraction and grouping of features

These data processing steps were carried out as recently described (Bueschl et al., 2013). In brief, for each resulting *m/z* cluster, the extracted ion chromatograms (EICs) of the monoisotopic ^{12}C (M) and the $^{13}C_6$ labelled (M+6) ions were evaluated for chromatographic peaks using the wavelet algorithm presented by Du, Kibbe, and Lin (2006). Only feature pairs consisting of corresponding chromatographic peaks for both the M and M+6 ions with similar retention times (within ±10 MS scans) and chromatographic peak shape (Pearson correlation coefficient $\geqslant 0.5$) were considered.

Subsequently, all monoisotopic ^{12}C features representing the same metabolite were grouped together by comparing the chromatographic peak shape of the feature of interest with those of all other features (Pearson correlation coefficient $\geqslant 0.85$) within a time interval of ±10 MS scans around the respective retention time. Feature groups were inspected to assign the frequently observed predefined adducts $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, $[M+K]^+$.

2.3.3. Construction of preliminary data matrix and target search for missing data

To compare relative feature levels between biological replicates and the two different treatments, the feature intensities (i.e. chromatographic peak areas) were organised in a matrix. Each row represented a distinct M/M+6 feature pair (ID of feature pair, ID of feature group, *m/z* value, retention time, type of ion species, if known). Samples were organised in rows, with the individual cells of the matrix containing peak areas of corresponding M and M+6 feature placed next to each other. Features potentially missed in any of the measurement files by the above mentioned strict parameter settings and/or low peak intensities were searched for in a targeted approach. This data matrix was further refined and extended by the weights of the individual samples and can be found in the supplemental material (Supplemental 2).

2.4. Statistical analysis

Prior to statistical analysis, for which R (R Development Core Team, 2012, V. 2.15.2) was used, all automatically acquired intensities were corrected by the weight of the respective berry skin sample and data still missing after targeted screening were filled by the arithmetic mean intensity value of the respective feature. For each data file, intensities (i.e. chromatographic peak areas) of all M and M+6 features within a particular feature group (i.e. metabolite) were summed, resulting in two intensity values (M and M+6) per metabolite and sample. Since groups of four grape berries were split in half with one set subjected to 25 °C and the other halves to 45 °C, a paired difference *t*-test was used to identify metabolites significantly differing (one-sided *t*-test, $p < 0.05$) between 25 °C and 45 °C treatments (Harris, 2010). For this, the summed areas of features from M and M+6 feature groups were used independently. Additionally, the relative difference [%] between 25 °C and 45 °C treatments was calculated for each significant feature group. This was achieved by calculating the mean difference of summed feature areas between the split berry halves relative to the condition with higher metabolite levels i.e. either the 25 °C or 45 °C treatment respectively.

3. Results and discussion

First, the acquired LC-HRMS raw chromatograms were inspected using TOPPView (Sturm & Kohlbacher, 2009). Highly abundant features containing M+6 patterns were clearly discernible in 2D chromatograms confirming the incorporation of the $^{13}C_6$-aromatic ring in newly formed metabolites and the formation of the expected isotopolog pattern (Fig. 1). In this respect, the native, ^{12}C isotopologs (M) of the respective metabolic features originate from non-labelled phenylalanine in the grape berries present prior to the experiment, while the isotopologs which carry the ^{13}C aromatic ring could have only been formed during the 120 h incubation with $^{13}C_6$ Phe and exhibited a typical *m/z* difference of 6.0201 u compared to their native isotopologs. This manual inspection also confirmed that no retention time shifts for the labelled isotopologs were present and that these isotopologs showed perfect co-elution with their native, non-labelled analogues. Using these highly abundant features, the enrichment of ^{13}C isotopes in the labelled phenylalanine tracer was calculated to be 98.8%. A comparison of the abundances observed for M and

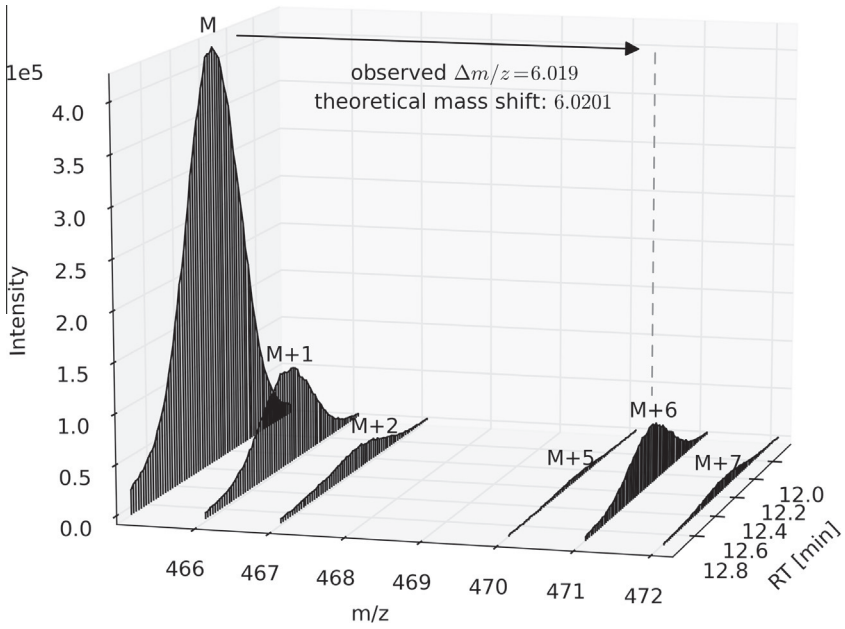


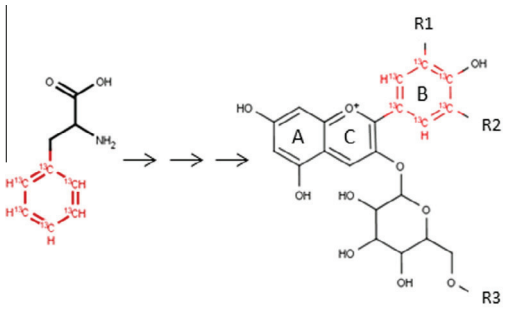
Fig. 1. Illustration of the isotopic pattern of metabolites incorporating one $^{13}C_6$ labelled aromatic ring of phenylalanine (feature pair #62; mono-isotopic ^{12}C *m/z*: 465.102; RT [min] 12.35). The M and M+1 ions are due to the natural isotopic composition of carbon (~99% ^{12}C and ~1% ^{13}C) in the non-labelled ion, while the corresponding M+5, M+6 and M+7 isotopologues can be attributed to the labelled analogue after incorporation of a $^{13}C_6$ aromatic ring (~1% ^{12}C and ~99% ^{13}C) and result from berry incubation with $^{13}C_6$ phenylalanine.

M+6 of manually selected feature pairs showed that M+6 ions were present at approximately 1/5th of the level of M ions, which can be explained by the fact that native, non-labelled metabolites were already accumulated in the grape berries prior to the experiment. These observations were used to perform the automatic search for M and M+6 ions containing either an unlabelled or an U-^{13}C labelled aromatic ring derived from the labelled phenylalanine (Fig. 2).

3.1. Molecular features containing M+6 label

Using the labelling assisted workflow described above, 110 analytical feature pairs were found across the 8 biological replicates of treatment and control samples (Supplemental 1). Each of these feature pairs consists of the monoisotopic ^{12}C (M) and its corresponding $^{13}C_6$ (M+6) ion species. All of these feature pairs can be assumed to represent metabolites, truly formed during the biological labelling experiment. Four extraction solvent blank samples were also measured. No labelling specific feature pairs were extracted from the data files of these blank samples. Some of the features found by the applied stable isotope-assisted approach occasionally showed low abundant chromatographic peaks in the blank samples with the experimental samples generally exhibiting 100-fold higher feature abundance. Both features of an M and M+6 feature pair can be characterised by a unique feature ID, *m/z* value, retention time and intensity value (chromatographic peak area). Assuming the ion species $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ and $[M+K]^+$ which belong to the most frequently formed adduct ions produced by electrospray ionisation, the molecular ion type could be assigned to 23 of the detected 110 features. Convolution of features into groups, representing single metabolites was based on retention time and peak shape similarity and resulted in 63 feature groups thus, each of which is representing a distinct metabolite. Barring false positives, this list (Supplemental 1) represents molecular ions that were labelled by $^{13}C_6$-aromatic ring. Metabolite ions carrying two labelled aromatic rings were also extracted by MetExtract, the assignment of the two aromatic rings was not performed automatically however. For example, feature ids 55 and 57 listed in Supplemental material 1 represent two distinct feature pairs. With the automated grouping they were however grouped to the same metabolite. A manual inspection then showed that the *m/z* values of these two feature pairs differed by exactly the same *m/z* value as the M and M+6 ions of each feature pair, confirming that these respective metabolites contained two labelled aromatic rings.

Using previously published *m/z* data of *V. vinifera* and anthocyanin degradation, possible identities were suggested for as many of the molecular features as possible (Bar-Akiva et al., 2010; Castillo-Munoz et al., 2009; Cavaliere et al., 2008; Mazzuca, Ferranti,



Anthocyanin	R1	R2	R3
Delphinidin	OH	OH	H
Cyanidin	OH	H	H
Petunidin	OCH_3	OH	H
Peonidin	OCH_3	H	H
Malvidin	OCH_3	OCH_3	H
O-acetyl-			
O-coumaryl-			

Fig. 2. Incorporation of $^{13}C_6$-labelled phenylalanine into different anthocyanin species.

Picariello, Chianese, & Addeo, 2005; Moss, Mao, Taylor, & Saucier, 2013; Pati, Liberatore, Gambacorta, Antonacci, & La Notte, 2009). Molecular features were then assessed for mass accuracy versus their possible identities. Accurate mass based annotations were confirmed by the presence of aglycone structures as in-source or product ion fragments when possible. In some cases, product ion mass spectra were compared to spectra obtained from the databases Metlin and Massbank. However, databases are still limited in their breadth so not all putatively identified metabolites were confirmed using spectral libraries.

Metabolites putatively identified include hydroxycinnamic acids, anthocyanins, flavonols, and stilbenes. While many of these compounds were expected, some of them were surprising. For example, stilbenes are known Phe derived metabolites in grapes, however their concentrations are typically quite low. However multiple resveratrol moieties were detected in berry samples. This may be due to the increased sensitivity of nanospray-ESI techniques (Yin & Killeen, 2007). Alternatively, resveratrol glucoside levels are known to increase under stress conditions such as water deficit (Deluc et al., 2010) as well as fungal infection (Roldan, Palacios, Caro, & Perez, 2003); we cannot say with certainty that, while sterile practices were used, there was no fungal pressure or water deficit stress on cultured berries. Furthermore, the act of culturing halved berries itself may induce resveratrol production.

3.2. Molecular features and feature groups significantly different between treatments

From the list of molecular features derived from $^{13}C_6$-Phe metabolism, a paired-comparison *t*-test was performed over the 8 biological replicates (for a full list of all features see Supplemental 2). Comparisons were made between the abundances the summed features of a group for both the ^{12}C and ^{13}C isotopolog. From the 2 comparisons made, multiple features or feature groups were significantly different between treatments for both the ^{12}C and ^{13}C isotopologs (Table 1).

In all, 6 feature groups of ^{12}C isotopologs and 10 feature groups of ^{13}C isotopologs differed significantly between 25 °C and 45 °C treatments (Fig. 3). Significant differences of individual, in-source fragment aglycone features should not be extrapolated to understand *in vivo* stability, however feature groups or features representing intact glycosides which showed lower abundances upon 45 °C treatment warrant consideration. For example, temperature reduced features such as myricetin and quercetin aglycones (data not shown) do not provide adequate support to interpret the thermal stability of their respective glycosides. Conversely, the reduction in tentatively identified resveratrol and resveratrol dimer feature groups provides strong evidence that heat treatment reduced their concentration in grape berries and may represent the first report of its thermal degradation *in vivo*. Moreover, earlier studies have also demonstrated that resveratrol is thermally labile as concentration levels in frozen berries dropped when exposed to cooking conditions (Lyons et al., 2003). Interestingly, Lyons et al. had found that the concentration of stilbene synthase (STS) as well as the transcription of its underlying gene had been positively influenced by the applied heat treatment. In another study, STS levels were transiently increased after just 8 h at 38 °C, however by 24 h, levels had returned to pre-treatment concentrations (Wang et al., 2008). Ultimately, with respect to our findings, any potential influence of STS at 45 °C was obviously not sufficient to counteract the reduction in intensity seen in the current experiment.

Among anthocyanins, the sum of all petunidin and malvidin glucoside associated features were significantly reduced after heat treatment. The reason for why only the feature groups of these two glucosides were affected is unclear. In physiological buffered extracts, earlier studies have reported that increased B-ring hydroxylation is associated with reduced anthocyanin stability (Woodward, Kroon, Cassidy, & Kay, 2009). However in grapes,

Table 1
Tentatively annotated feature groups of ^{12}C and ^{13}C isotopologs that varied significantly due to temperature treatment. Bold assignments represent the single compound responsible for all features within a given feature group. Annotations were made based on accurate mass, MS/MS confirmation, and known presence in grape berries.

MetGroupID	Metabolite	Representative features (monoisotopic ^{12}C *m/z*)	RT [min]	*m/z* experimental	*m/z* calculated	Mass error (ppm)	MS^2	MS^2 database	^{13}C mean difference [%]	^{12}C mean difference [%]
63	Unknown 1		10.7	611.1615					−32*	8
82	Unknown 2		13.1	669.1642					−27*	2
24	**Petunidin-3-glucoside**	Petunidin-3-glucoside	13.5	479.1184	479.1189[a]	1.21	Y	Massbank	−16*	5
		Petunidin	13.5	317.0651	317.0661[a]	3.31				
		Petunidin H_2O	13.5	335.0761						
		Petunidin-3-glucoside H_2O	13.5	497.1274						
31	**Malvidin-3-glucoside**	Malvidin-3-glucoside	14.5	493.1341	493.1346[a]	0.98	Y	Massbank	−12*	1
		Malvidin-3-glucoside H_2O	14.5	511.1442	511.1451[a]	1.91				
		Malvidin	14.5	331.0812	331.0817[a]	1.74				
		Malvidin H_2O	14.5	349.0915	349.0923[a]	2.55				
		Syringetin	14.5	347.0761	347.0761[b]	0.11				
36	Unknown 3		15.7	434.2021					2	22*
37	Unknown 4		16.2	434.2008					13*	13*
39	Unknown 5		16.8	448.2168					9	22*
13	Unknown 6		19.6	243.0652					−26*	−1
58	Unknown 7		20.2	585.1704					13	29*
11	**Resveratrol**	Resveratrol	21.7	229.0860	229.0859[b]	−0.54	Y	Massbank	−50*	−44*
57	Unknown 8		21.7	565.1916					61*	5
48	**Scirpusin A**	Scirpusin A	22.1	471.1432	471.1438[b]	1.18			−30*	−30*
		Scirpusin A [M+6]	22.1	477.1628	477.1640[b]	2.60				
42	**Resveratrol dimer**	Resveratrol dimer	22.6	455.1484	455.1489[b]	1.04			−28*	−19
		Resveratrol dimer [M+6]	22.6	461.1681	461.1690[b]	1.86				

[a] Molecular feature was present as $[M]^+$ ion.
[b] Molecular feature was present as $[M+H]^+$ ion.
* Statistically significant difference between 25 °C and 45 °C at $p = 0.05$.

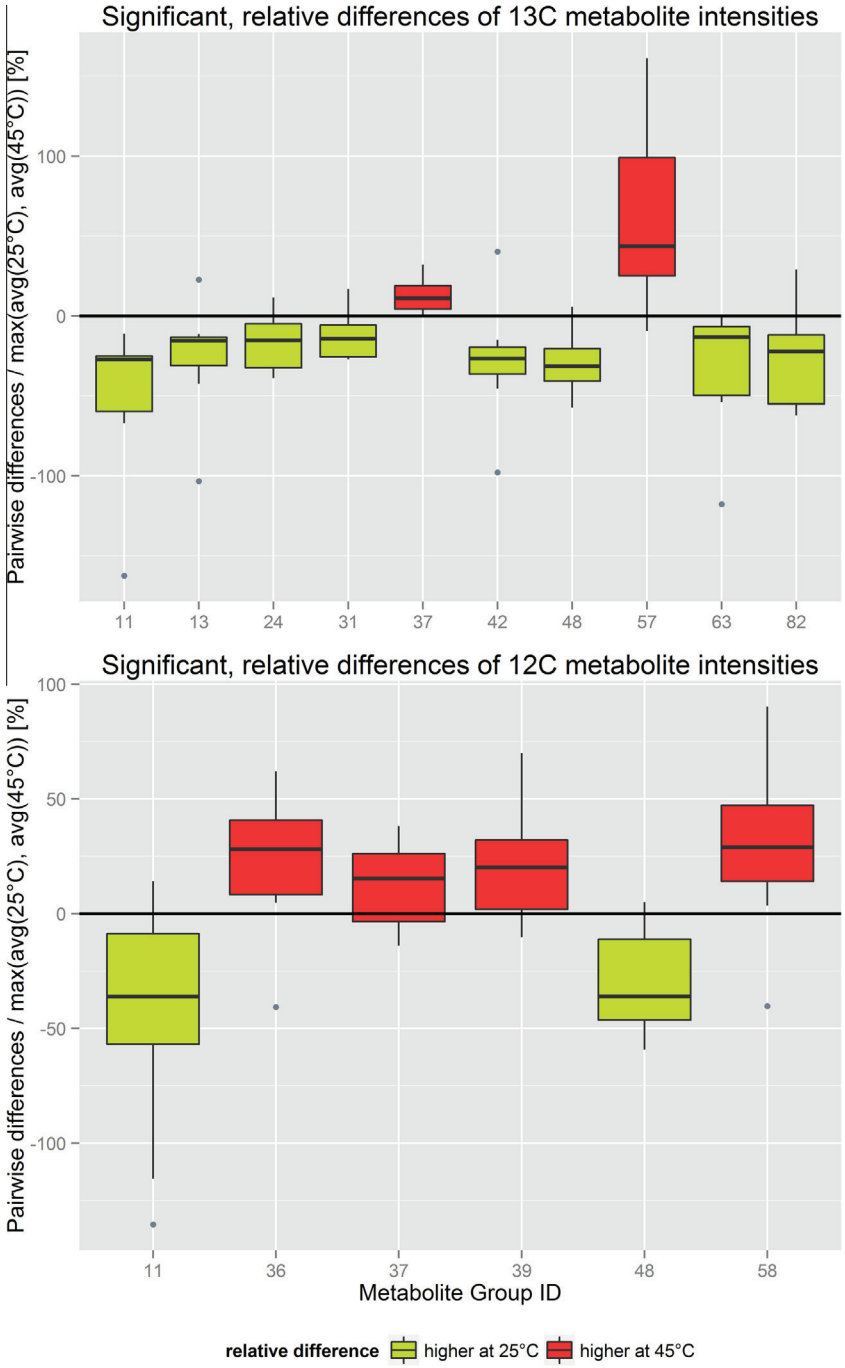


Fig. 3. Metabolites significantly differing between heat treated and control samples. Each boxplot represents the distribution of relative differences in metabolite signal abundances for corresponding berry halves. Metabolite Group IDs refer to those shown in Table 1 (column MetGroupID) and Supplemental 2 (column MetGroupID). The upper and lower hinge represent the 25th and 75th percentiles (inter-quartile range; IQR) and the whiskers extend to a maximum of 1.5*IQR. Grey dots show extreme values.

increased temperatures are associated with increased proportions of hydroxylated and methylated anthocyanin B-rings (Tarara, Lee, Spayd, & Scagel, 2008). None of the acetylated or coumarylated anthocyanin related feature groups were significantly influenced by the temperature treatment in our study. This observation is similar to previous research, as acylated and coumarylated anthocyanins are known to be more stable than their biosynthetic precursor glucosides (Mori et al., 2007).

The decrease in anthocyanin levels at elevated temperature is well known and has been shown in many plant species including grapes (Mori et al., 2007) and therefore, the chemical non-enzymatic degradation of these compounds due to thermal processing is of concern for food manufacturers (Patras, Brunton, O'Donnell, & Tiwari, 2010). In apple for example, warm weather conditions have been shown to reduce the concentration of anthocyanins and expression of the transcription factor *MYB10*, known for regulation of anthocyanin production (Lin-Wang et al., 2011). In overexpressing *A. thaliana* mutants, low light and 30 °C conditions reduced anthocyanin concentrations despite overexpression of the *PAP1* MYB transcription factor, suggesting active catabolism of anthocyanins (Rowan et al., 2009). At the same time however, no catabolic genes involved in anthocyanin degradation had been found to be differentially expressed, suggesting that chemical, non-enzymatic degradation may contribute to the loss of anthocyanins. As the presented approach involves simultaneous automated detection of pairs of labelled and unlabelled

isotopologs, it facilitates the identification of all "real" biology derived molecular features, including those representing putative degradation products. The levels of five feature groups increased significantly when exposed to 45 °C: unknowns 3, 4, 5, 7, and 8 (Table 1) and thus can be considered candidate metabolites originating from anthocyanin or other flavonoid degradation. Unfortunately, meaningful MS/MS spectra for these features were not achieved due to their low abundance in the investigated samples. Further work to investigate the molecular structure of these potential flavonoid degradation products is necessary.

4. Conclusion

The presented study demonstrates the high potential of stable isotopic tracer assisted approaches for the global untargeted profiling of tracer derived metabolites in complex biological samples. While for conventional untargeted metabolomics workflows, which do not make use of isotopic labelling, data evaluation is greatly complicated by the large majority of non-tracer related metabolites, the presented approach directly enables monitoring substrate-associated, discrete biochemical pathways. Furthermore, the specificity of this approach allows direct interpretation of changes in those metabolic processes with the involvement of tracer containing metabolites, without the influence of tangential biochemical processes which would cloud statistical interpretation. In the current work, which probed the metabolic response to heat treatment in grape berries and filtered the results for compounds related to the phenylpropanoid pathway, 13 feature groups showed significant differences between 45 °C and 25 °C degree treatments *in vivo*. Only a few molecular features were found to be present at significantly higher levels in the 45 °C treatment. These compounds present good candidates for degradation products of anthocyanins and the detailed investigation of these metabolites will present an opportunity to possibly assign anthocyanin derived degradation products or novel phenylalanine derivatives. Moreover, future studies using MS/MS will help in identifying the biochemical metabolites represented by the molecular features that may result from anthocyanin degradation.

Acknowledgements

The authors would like to thank American Vineyard Foundation for financial support.

A.W.C. was supported by NIH 1U24DK097154 (PI Fiehn, Oliver) and graduate student fellowships, including the Mario P. Tribuno Memorial Research Fellowship, Wine Spectator Scholarship, Horace O. Lanza Scholarship, Adolf L. & Richie C. Heck Research Scholarship, Leon D. Adams Research Scholarship, and Curtis J. Alley Memorial Research Scholarship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2014.06.002.

References

Bar-Akiva, A., Ovadia, R., Rogachev, I., Bar-Or, C., Bar, E., Freiman, Z., et al. (2010). Metabolic networking in Brunfelsia calycina petals after flower opening. *Journal of Experimental Botany, 61*(5), 1393–1403.

Boatright, J., Negre, F., Chen, X., Kish, C., Wood, B., Peel, G., et al. (2004). Understanding in vivo benzenoid metabolism in petunia petal tissue. *Plant Physiology, 135*(4), 1993.

Bueschl, C., Kluger, B., Berthiller, F., Lirk, G., Winkler, S., Krska, R., et al. (2012). MetExtract: A new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. *Bioinformatics, 28*(5), 736–738.

Bueschl, C., Kluger, B. L., Marc, Adam, G., Wiesenberger, G., Maschietto, V., et al. (2013). A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research. *Metabolomics*. http://dx.doi.org/10.1007/s11306-013-0611-0.

Castillo-Munoz, N., Gomez-Alonso, S., Garcia-Romero, E., Gomez, M. V., Velders, A. H., & Hermosin-Gutierrez, I. (2009). Flavonol 3-O-glycosides series of *Vitis vinifera* Cv. Petit Verdot Red Wine Grapes. *Journal of Agricultural and Food Chemistry, 57*(1), 209–219.

Cavaliere, C., Foglia, P., Gubbiotti, R., Sacchetti, P., Samperi, R., & Lagana, A. (2008). Rapid-resolution liquid chromatography/mass spectrometry for determination and quantitation of polyphenols in grape berries. *Rapid Communications in Mass Spectrometry, 22*(20), 3089–3099.

Charron, C., Britz, S., Mirecki, R., Harrison, D., Clevidence, B., & Novotny, J. (2008). Isotopic labeling of red cabbage anthocyanins with atmospheric $^{13}CO_2$. *Journal of the American Society for Horticultural Science, 133*(3), 351–359.

Deluc, L. G., Decendit, A., Papastamoulis, Y., Mérillon, J.-M., Cushman, J. C., & Cramer, G. R. (2010). Water deficit increases stilbene metabolism in Cabernet Sauvignon Berries. *Journal of Agricultural and Food Chemistry, 59*(1), 289–297.

Du, P., Kibbe, W. A., & Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics, 22*(17), 2059–2065.

Ferrer-Gallego, R., Hernández-Hierro, J. M., Rivas-Gonzalo, J. C., & Escribano-Bailón, M. T. (2012). Influence of climatic conditions on the phenolic composition of *Vitis vinifera* L. cv. Graciano. *Analytica Chimica Acta, 732*, 73–77.

Forester, S. C., & Waterhouse, A. L. (2010). Gut metabolites of anthocyanins, gallic acid, 3-O-methylgallic acid, and 2,4,6-trihydroxybenzaldehyde, inhibit cell proliferation of Caco-2 cells. *Journal of Agricultural and Food Chemistry, 58*(9), 5320–5327.

Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H. M., Hesse, H., et al. (2011). Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. *The Plant Journal, 68*(2), 364–376.

Harris, D. C. (2010). *Quantitative chemical analysis*. Macmillan.

Hegeman, A. D., Schulte, C. F., Cui, Q., Lewis, I. A., Huttlin, E. L., Eghbalnia, H., et al. (2007). Stable isotope assisted assignment of elemental compositions for metabolomics. *Analytical Chemistry, 79*(18), 6912–6921.

Heinzle, E., Matsuda, F., Miyagawa, H., Wakasa, K., & Nishioka, T. (2007). Estimation of metabolic fluxes, expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation. *The Plant Journal, 50*(1), 176–187.

Kluger, B., Bueschl, C., Lemmens, M., Berthiller, F., Häubl, G., Jaunecker, G., et al. (2012). Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat. *Analytical and Bioanalytical Chemistry*, 1–6.

Lin-Wang, K. U. I., Micheletti, D., Palmer, J., Volz, R., Lozano, L., Espley, R., et al. (2011). High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. *Plant, Cell & Environment, 34*(7), 1176–1190.

Lyons, M. M., Yu, C., Toma, R. B., Cho, S. Y., Reiboldt, W., Lee, J., et al. (2003). Resveratrol in raw and baked blueberries and bilberries. *Journal of Agricultural and Food Chemistry, 51*(20), 5867–5870.

Matsuda, F., Morino, K., Miyashita, M., & Miyagawa, H. (2003). Metabolic flux analysis of the phenylpropanoid pathway in wound-healing potato tuber tissue using stable isotope-labeled tracer and LC–MS spectroscopy. *Plant and Cell Physiology, 44*(5), 510–517.

Mazzuca, P., Ferranti, P., Picariello, G., Chianese, L., & Addeo, F. (2005). Mass spectrometry in the study of anthocyanins and their derivatives: differentiation of *Vitis vinifera* and hybrid grapes by liquid chromatography electrospray ionization mass spectrometry and tandem mass spectrometry. *Journal of Mass Spectrometry, 40*(1), 83–90.

Mori, K., Goto-Yamamoto, N., Kitayama, M., & Hashizume, K. (2007). Loss of anthocyanins in red-wine grape under high temperature. *Journal of Experimental Botany, 58*(8), 1935–1945.

Moss, R., Mao, Q., Taylor, D., & Saucier, C. (2013). Investigation of monomeric and oligomeric wine stilbenoids in red wines by ultra-high-performance liquid chromatography/electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry, 27*(16), 1815–1827.

Neish, A. C. (1960). Biosynthetic pathways of aromatic compounds. *Annual Review of Plant Physiology and Plant Molecular Biology, 11*, 55–80.

Pati, S., Liberatore, M., Gambacorta, G., Antonacci, D., & La Notte, E. (2009). Rapid screening for anthocyanins and anthocyanin dimers in crude grape extracts by high performance liquid chromatography coupled with diode array detection and tandem mass spectrometry. *Journal of Chromatography A, 1216*(18), 3864–3868.

Patras, A., Brunton, N. P., O'Donnell, C., & Tiwari, B. (2010). Effect of thermal processing on anthocyanin stability in foods; mechanisms and kinetics of degradation. *Trends in Food Science & Technology, 21*(1), 3–11.

Roldan, A., Palacios, V., Caro, I., & Perez, L. (2003). Resveratrol content of Palomino fino grapes: Influence of vintage and fungal infection. *Journal of Agricultural and Food Chemistry, 51*(5), 1464–1468.

Rowan, D., Cao, M., Lin-Wang, K., Cooney, J., Jensen, D., Austin, P., et al. (2009). Environmental regulation of leaf colour in red 35S: PAP1 *Arabidopsis thaliana*. *New Phytologist, 182*(1), 102–115.

Sinilal, B., Ovadia, R., Nissim-Levi, A., Perl, A., Carmeli-Weissberg, M., & Oren-Shamir, M. (2011). Increased accumulation and decreased catabolism of

anthocyanins in red grape cell suspension culture following magnesium treatment. *Planta, 234*(1), 61–71.
Somers, C. (1998). *The wine spectrum*. Adelaide, Australia: Hyde Park Press.
Sturm, M., & Kohlbacher, O. (2009). TOPPView: An open-source viewer for mass spectrometry data. *Journal of Proteome Research, 8*(7), 3760–3763.
Tarara, J. M., Lee, J. M., Spayd, S. E., & Scagel, C. F. (2008). Berry temperature and solar radiation alter acylation, proportion, and concentration of anthocyanin in Merlot grapes. *American Journal of Enology and Viticulture, 59*(3), 235–247.
Wang, W., Wan, S.-B., Zhang, P., Wang, H.-L., Zhan, J.-C., & Huang, W.-D. (2008). Prokaryotic expression, polyclonal antibody preparation of the stilbene synthase gene from grape berry and its different expression in fruit development and under heat acclimation. *Plant Physiology and Biochemistry, 46*(12), 1085–1092.
Winkler, A., Cook, J., Kliewer, W., & Lider, L. (1962). *Development and composition of grapes*. General Viticulture: University of California Press, Berkeley. 4.
Woodward, G., Kroon, P., Cassidy, A., & Kay, C. (2009). Anthocyanin stability and recovery: Implications for the analysis of clinical and experimental samples. *Journal of Agricultural and Food Chemistry, 57*(12), 5271–5278.
Yin, H., & Killeen, K. (2007). The fundamental aspects and applications of Agilent HPLC-Chip. *Journal of Separation Science, 30*(10), 1427–1434.

Publication #5: Untargeted profiling of tracer derived metabolites using stable isotopic labelling and fast polarity switching LC-ESI-HRMS

Bernhard Kluger†, **Christoph Bueschl**†, Nora KN. Neumann, Romana Stückler, Maria Doppler, Alexander Chassy, Andrew Waterhouse, Justyna Rechthaler, Niklas Kampleitner, Gerhard G. Thallinger, Gerhard Adam, Rudolf Krska, Rainer Schuhmacher

Analytical Chemistry. 2014;86(23):11533-11537

Contributions of the presenting author: The author of this thesis implemented the functionality to search for SIL-derived biotransformation products, extended MetExtract to support fast polarity switching derived LC-HRMS data, and performed the automated data processing of the acquired LC-HRMS data. Additionally, the author of this thesis performed the preliminary statistical investigation of the DON-stressed and control samples. Together with Bernhard Kluger, the author of this thesis wrote the text of the publication. Since this publication heavily depends on the interdisciplinary research between analytical chemistry and bioinformatics, the authors agreed that this publication was realised with shared first author option.

Supplementary material: Certain parts of the manuscript were transferred to supplementary material S1, which is not included in this work but can be accessed online at the journal's homepage (https://dx.doi.org/10.1021/ac503290j).

† Authors have contributed equally to this work

analytical chemistry

Technical Note

pubs.acs.org/ac

Untargeted Profiling of Tracer-Derived Metabolites Using Stable Isotopic Labeling and Fast Polarity-Switching LC–ESI-HRMS

Bernhard Kluger,[†,¶] Christoph Bueschl,[†,¶] Nora Neumann,[†] Romana Stückler,[‡] Maria Doppler,[†] Alexander W. Chassy,[§] Andrew L. Waterhouse,[§] Justyna Rechthaler,[∥] Niklas Kampleitner,[∥] Gerhard G. Thallinger,[⊥,#] Gerhard Adam,[‡] Rudolf Krska,[†] and Rainer Schuhmacher*[,†]

[†]Center for Analytical Chemistry, Department for Agrobiotechnology (IFA-Tulln), University of Natural Resources and Life Sciences Vienna (BOKU), Konrad-Lorenz-Strasse 20, 3430 Tulln, Austria

[‡]Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences Vienna (BOKU), Konrad-Lorenz-Strasse 24, 3430 Tulln, Austria

[§]Department of Viticulture and Enology, University of California Davis, Davis, California 95616, United States

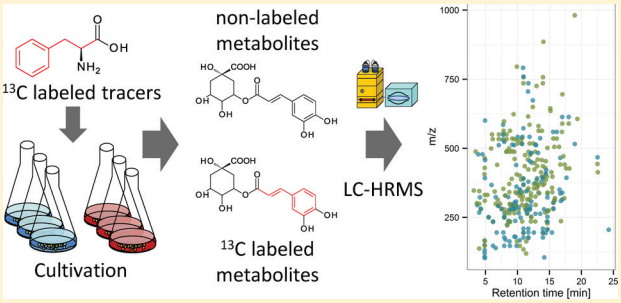
[∥]University of Applied Sciences Wr. Neustadt, Degree Programme Biotechnical Processes (FHWN-Tulln), Konrad Lorenz Strasse 10, 3430 Tulln, Austria

[⊥]Bioinformatics Group, Institute for Knowledge Discovery, Graz University of Technology, Petersgasse 14, 8010, Graz, Austria

[#]BioTechMed OMICS Center Graz, Stiftingtalstraße 24, 8010, Graz, Austria

Supporting Information

ABSTRACT: An untargeted metabolomics workflow for the detection of metabolites derived from endogenous or exogenous tracer substances is presented. To this end, a recently developed stable isotope-assisted LC–HRMS-based metabolomics workflow for the global annotation of biological samples has been further developed and extended. For untargeted detection of metabolites arising from labeled tracer substances, isotope pattern recognition has been adjusted to account for nonlabeled moieties conjugated to the native and labeled tracer molecules. Furthermore, the workflow has been extended by (i) an optional ion intensity ratio check, (ii) the automated combination of positive and negative ionization mode mass spectra derived from fast polarity switching, and (iii) metabolic feature annotation. These extensions enable the automated, unbiased, and global detection of tracer-derived metabolites in complex biological samples. The workflow is demonstrated with the metabolism of $^{13}C_9$-phenylalanine in wheat cell suspension cultures in the presence of the mycotoxin deoxynivalenol (DON). In total, 341 metabolic features (150 in positive and 191 in negative ionization mode) corresponding to 139 metabolites were detected. The benefit of fast polarity switching was evident, with 32 and 58 of these metabolites having exclusively been detected in the positive and negative modes, respectively. Moreover, for 19 of the remaining 49 phenylalanine-derived metabolites, the assignment of ion species and, thus, molecular weight was possible only by the use of complementary features of the two ion polarity modes. Statistical evaluation showed that treatment with DON increased or decreased the abundances of many detected metabolites.



Untargeted metabolomics approaches probe the entire metabolic space of a biological system (e.g., cells or whole organism). This can be realized by trying to measure as many metabolites as possible or alternatively by searching for those metabolites that arise from either exogenous or endogenous substances such as toxins, drugs, or sugars and amino acids, respectively. The screening of such metabolites in LC–HRMS data is rather straightforward when performed in (a) a targeted way with positive lists of putative biotransformation products (e.g., Levsen et al.,[1] Sandermann[2]). In contrast, untargeted approaches are usually more challenging and aim at the detection of known and unknown metabolic products by (b) background subtraction and/or statistical investigation (e.g., Zhang et al.[3]) or (c) isotopic labeling, including stable isotopic labeling (SIL)-assisted approaches. Although the search according to approach a is limited to the subset of predicted, putative tracer derived conjugates (e.g., sugars, amino acids, small peptides) or degradation products known from literature and previous approaches, approach b also enables detecting previously unknown metabolites, but also requires more sophisticated data processing than approach a. Furthermore, the latter approach is prone to detect non-tracer-related metabolites, significantly differing between the investigated

Received: September 2, 2014
Accepted: November 5, 2014
Published: November 5, 2014

   dx.doi.org/10.1021/ac503290j | Anal. Chem. 2014, 86, 11533–11537

sample types. In contrast, approach c provides an easy way to detect both known and unknown tracer-derived metabolites and has the advantage over both a and b that the detected metabolites can be linked to the studied tracer substance (e.g., Baillie,[4] Iglesias et al.[5]).

To avoid the use of radioisotopes, SIL-assisted metabolism studies use stable isotope (e.g., ^{13}C, ^{15}N, or ^{34}S)-enriched tracers and assume that biological systems metabolize native and labeled variants of a supplied tracer nearly equally.[6] Cabaret and colleagues[7] studied U-^{13}C sterigmatocystin in porcine tracheal epithelial cells, and Li and colleagues[8] utilized deuterium labeling together with a principal component analysis guided approach to detect novel metabolites of the drug tempol.

For GC/MS-based, untargeted tracer metabolism studies, Hiller and colleagues[9] presented the NTFD (nontargeted tracer fate detection) algorithm, which detects changes and metabolic fluxes derived from labeled tracers in the primary metabolome. For LC−HRMS-based tracer metabolization approaches, several software tools designed for the untargeted detection and analysis of isotope patterns of metabolites derived from native and partly isotopically labeled tracers are available (e.g., mzMatch-ISO,[10] X^{13}CMS[11]). However, to the best of our knowledge, no tools for the automated global and highly selective detection of only those metabolites derived from native and labeled tracers with nonoverlapping isotope patterns are currently available.

Thus, a LC−HRMS-based workflow for the unbiased detection of known and unknown metabolites derived from U-^{13}C-SIL guided tracer metabolism was developed. It is based on our recently published workflow for the detection of metabolic features derived from native and fully labeled biological samples,[12] which has been further developed to support fast polarity switching and automated annotation of metabolic features of the detected metabolites. In contrast, to the currently existing workflows, such as mzMatch-ISO or fluxomics applications, which have been designed to detect shifts of relative abundances in native isotope patterns, the presented approach requires distinct, nonoverlapping isotope patterns and is capable of detecting metabolites for which the native and labeled analogues differ by ≥4 u. Therefore, it is mainly suited to study the secondary metabolism of a biological system of interest rather than to support the elucidation of primary metabolism. Moreover, the use of nonoverlapping isotope patterns enables determining the exact number of incorporated carbon atoms of the employed tracer in the respective biotransformation product and thus improves sum formula and metabolite annotation. At a less advanced stage, the presented concept has already been used successfully to study the metabolic fate of the mycotoxin deoxynivalenol (DON) in wheat (*Triticum aestivum*, Kluger et al.[13]) and the fate of the aromatic amino acid phenylalanine in grape berries (*Vitis vinifera*, Chassy et al.[14]). Here, our approach is presented with the metabolism of the endogenous amino acid phenylalanine in wheat cell suspension cultures in the presence or absence of the *Fusarium* virulence factor DON. Phenylalanine was chosen as a tracer because it serves as precursor for the biosynthesis of hydroxycinnamic acids, phenylpropanoids, and flavonoids in plants, many of which are known to be involved in the defense against fungal pathogens such as *Fusarium*.[15]

MATERIALS AND METHODS

Chemicals. Acetonitrile (ACN, HiPerSolv Chromanorm, HPLC gradient grade) was purchased from VWR (Vienna, Austria), methanol (MeOH, LiChrosolv, LC gradient grade) was purchased from Merck (Darmstadt, Germany), and formic acid (FA, MS grade) was obtained from Sigma-Aldrich (Vienna, Austria). Water was purified successively by reverse osmosis and an ELGA Purelab Ultra-AN-MK2 system (Veolia Water, Vienna, Austria). U-$^{13}C_9$ phenylalanine (U-$^{13}C_9$ Phe; 99.1% ^{13}C) was purchased from Euriso-top (Saarbrücken, Germany).

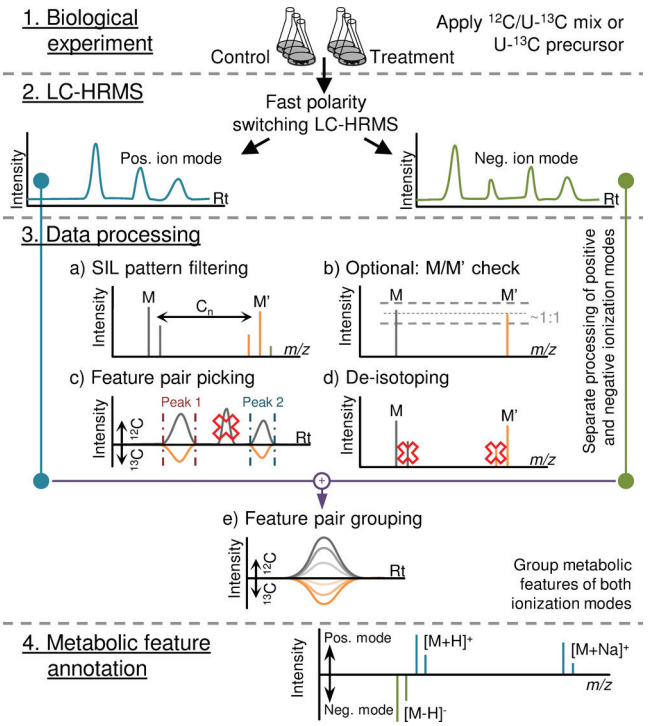


Figure 1. Illustration of the workflow for SIL-assisted tracer metabolism studies, including LC−ESI-HRMS fast polarity switching.

Biological Experiment (Figure 1, Step 1). Aliquots (3.6 mL) of *T. aestivum* (*Tae*) cell suspension cultures in B5 medium (Supporting Information S1.1) were incubated with 400 μL of aqueous solutions differing in composition according to the three tested conditions (3 replicates per condition). Thus, each culture sample resulted in a final volume of 4 mL. For the condition "control", 200 μL of U-$^{13}C_9$ Phe stock solution (final concentration in culture: 25 mg/L) and 200 μL of H_2O dist. were added to the culture. For cocultivation with DON (condition "treatment"), 200 μL of U-$^{13}C_9$ Phe stock solution (final concentration in culture: 25 mg/L) and 200 μL of DON stock solution (final concentration in culture: 90 mg/L) were added. For the condition "blank" 400 μL of H_2O dist. was added to 3.6 mL of cell suspension culture. All *Tae* cell suspension cultures were grown in 25 mL Erlenmeyer flasks for 8 days at 20 °C in the light with shaking (100 rpm). In addition, two medium blanks without cell suspension cultures were prepared in parallel.

Sample Preparation. After cultivation, 2 mL of each sample was transferred into savelock Eppendorf tubes and centrifuged for 5 min at 5000*g*. The weight of each cell pellet was determined, and 350 μL MeOH and 2 × 5 mm steel balls

dx.doi.org/10.1021/ac503290j | *Anal. Chem.* 2014, 86, 11533−11537

were added prior to wet milling with a ball mill (MM 301 Retsch, Haan, Germany). Samples were homogenized for 2 min at 30 Hz, and the suspension was transferred to new Eppendorf tubes and centrifuged for 5 min at 20 000*g*. From the resulting extract, 200 μL was mixed with 200 μL of H_2O dist. and centrifuged for 10 min at 20 000*g* again. An aliquot of 300 μL of this sample solution was transferred to an HPLC vial and stored at −80 °C until further analysis.

LC−HRMS Analysis (Figure 1, Step 2). All samples were analyzed on a UHPLC system (UltiMate 3000, Dionex) coupled to an Orbitrap Exactive Plus (Thermo Fisher Scientific) equipped with a heated electrospray ionization (ESI) source. A Dionex autosampler was used for the injection of 10 μL per sample for chromatographic separation at 25 °C on a reversed-phase XBridge C18 150 × 2.1 mm i.d., 3.5 μm column (Waters, Milford, MA, USA) preceded by a C18 4 × 3 mm i.d. security cartridge (Phenomenex, Torrance, CA, USA). At a constant flow rate of 250 μL/min, a linear gradient program with water containing 0.1% FA (v/v) (eluent A) and MeOH containing 0.1% FA (v/v) was employed;[12] the initial mobile phase composition (10% eluent B) was held constant for 2 min, followed by a linear gradient to 100% eluent B within 30 min. After a hold time of 5 min, the column was re-equilibrated for 8 min at 10% eluent B.

The heated ESI interface was operated in fast polarity-switching mode using the following settings for both polarities: sheath gas, 50 au; auxiliary gas, 5 au; capillary voltage, 3 kV; capillary temperature, 350 °C. FT-Orbitrap was operated in profile mode (scan range, m/z 100−1000) with a resolving power of 70 000 fwhm (at m/z 200) and automatic gain control setting of 3×10^6 with a maximum injection time of 200 ms.

SIL-Assisted Data Processing (Figure 1, Steps 3 and 4). The SIL-assisted data processing for metabolite detection in full metabolome labeling experiments described earlier[12] was extended to support the detection of tracer-derived metabolites. It is part of a software package that is currently under development and will comprise three different data processing workflows for SIL-assisted metabolomics approaches. In the meantime, the software module facilitating data processing according to the presented workflow is accessible via the corresponding author.

Each of the following data processing steps for metabolic feature detection is carried out independently for the positive and negative ionization mode: First, every MS scan is inspected for pairs of two mass peaks, M, which corresponds to a native metabolization product, and M′, which denotes the same metabolization product but contains the labeled tracer molecule or a part of it, with an m/z difference proportional to the number of tracer-derived heavy isotope atoms (here, ^{13}C) present in the labeled metabolite ion (step 3a). A peak pair is accepted if the observed mass difference is within a preset maximum mass tolerance window of that calculated for the algorithm-predicted number of heavy isotopes. Optionally, for exogenous tracers, the intensity ratio I_M:$I_{M'}$ is compared with that of the concentration ratio of native and labeled tracer used for sample incubation (step 3b).

Next, the observed isotopolog ratio $I_{M'-1}/I_{M'}$ is compared with its theoretical ratio expected from the number of labeling isotopes contained in the inspected metabolite ion (step 3a). The corrected intensity ratio of the isotopologs $I_{M+1}/I_M - I_{M'+1}/I_{M'}$, which accounts for any carbon atom in the nonlabeled moiety, is compared with its theoretical ratio as determined for the number of labeling isotopes of the respective M and M′ ion pair. All intensity ratio checks are passed if the relative deviation between the expected and the observed ion intensity ratios are within preset error windows. Then, all such detected M and M′ pairs from different scans are combined with hierarchical clustering using the assigned number of heavy isotopes per metabolite ion and the m/z value of M. Clusters showing a maximum relative mass deviation between their highest and lowest m/z value of less than a preset threshold in parts per million are not further split.

Next, for each M and M′ ion cluster, chromatographic peaks in the corresponding $^{12}C/^{13}C$ EICs are extracted with the wavelet algorithm of Du et al.[16](step 3c). Only those EIC peaks that are found for both ^{12}C and corresponding ^{13}C m/z traces closely coelute and have a high Pearson correlation coefficient remain for further data processing. Incorrectly detected M and M′ pairs that originate from carbon isotopologs (e.g., M + 1 instead of M or M′ − 1 instead of M′) are removed from the data (step 3d). Correctly assigned chromatographic peak pairs are finally listed as metabolic features, each consisting of a ^{12}C monoisotopic m/z for M, a retention time (t_R), feature areas determined for the EIC peaks of M and M′, and the number of heavy isotopes originating from the tracer.

Following metabolic feature detection, all features found in positive or negative ionization mode are combined across both ionization polarities to generate feature groups, each of which represents a distinct metabolite (step 3e). To this end, the Pearson correlation coefficient is calculated pairwise for closely coeluting metabolic features. All metabolic features with a correlation coefficient above a preset threshold are put into the same feature group.

Subsequently, each determined feature group is annotated (step 4). For this, all features of a feature group are inspected pairwise for m/z differences corresponding to predicted ion species frequently found for the respective ionization mode. For metabolic features without a valid adduct pairing, neutral losses are calculated according to the Seven Golden Rules.[17]

LC−HRMS data derived from wheat cell suspension culture samples were processed as described above with the following parameter settings: (step 3a) Isotopic carbon enrichment, 98.9% ^{12}C; 99.1% $^{13}C_9$-Phe; Δ m/z $^{12}C/^{13}C$, 1.00335 u; atom counts, 6−9; minimum intensity threshold of putative M and M′ signals, 50 000 counts; maximum mass deviation between M and putatively corresponding M′ signals, ±3 ppm; maximum isotopolog ratio error, 20%. (step 3c) EIC m/z width, ± 5 ppm; minimum correlation coefficients between EIC peaks of M and M′, 0.75 for peak picking and 0.9 for feature grouping (step 3e). (step 4) Adducts used for feature annotation, $[M + H]^+$, $[M + Na]^+$, $[M + NH_4]^+$, $[M + K]^+$, $[M - e]^+$, $[M - H]^-$, $[M + FA - H]^-$, $[M + Na - 2H]^-$, $[M + Cl]^-$, $[M + K - 2H]^-$, $[M + Br]^-$, $[M - e - 2H]^-$. The accurate m/z values of all detected metabolites were searched against a custom wheat metabolite database containing 1145 entries (max m/z difference ≤ 5 ppm). Statistical evaluation of the experiment is described in Supporting Information S1.3.

RESULTS AND DISCUSSION

In a biological system, native and ^{13}C-enriched substances are metabolized by the same biological transformations and to a nearly equal extent.[6] As a result, all metabolites derived from a mixture of native and ^{13}C-labeled tracer contain either the entire native or the enriched tracer or just a part of it. In LC−HRMS, native metabolites and their corresponding ^{13}C isotopologs perfectly coelute with highly similar chromato-

dx.doi.org/10.1021/ac503290j | Anal. Chem. 2014, 86, 11533−11537

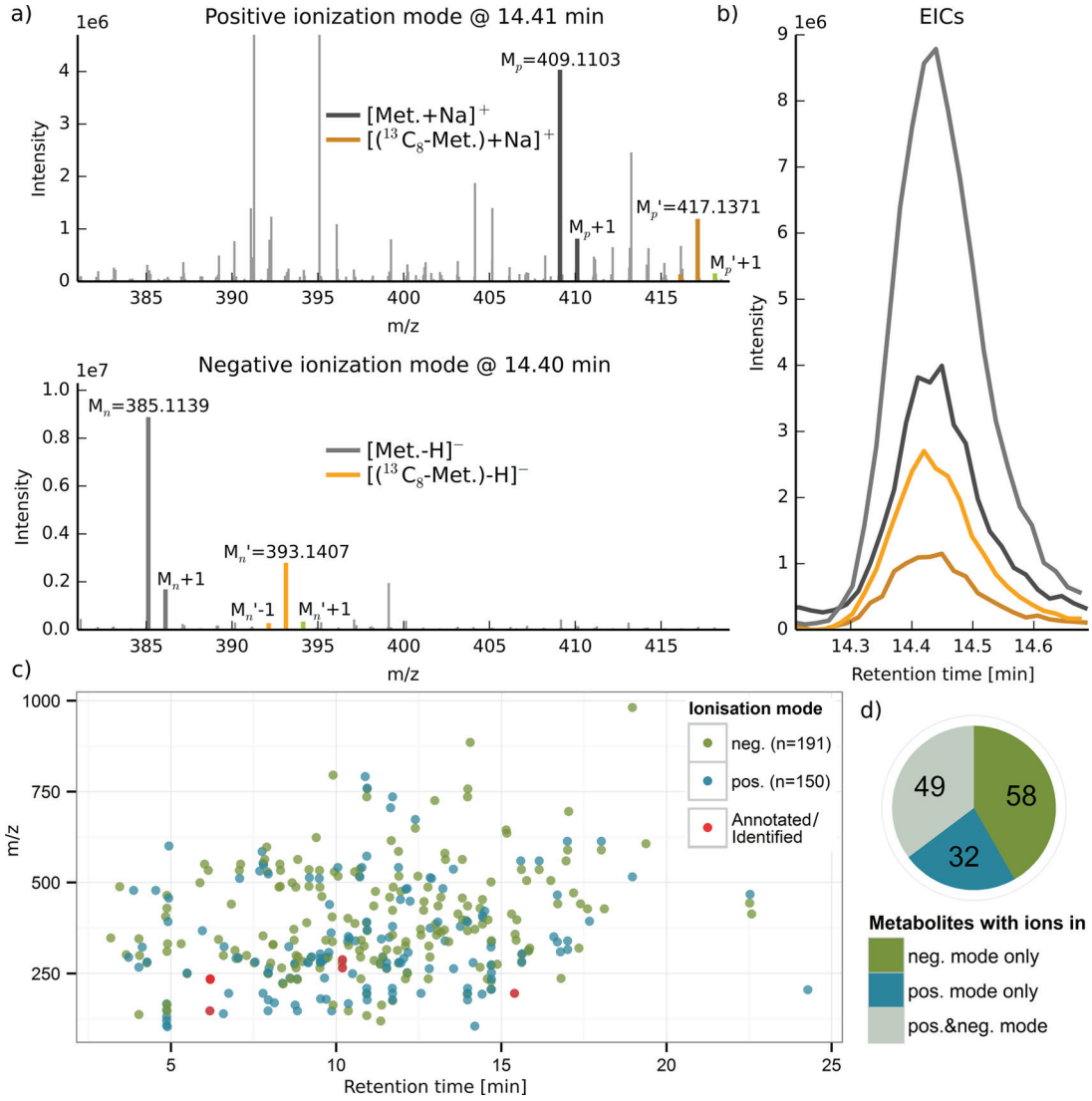


Figure 2. (a, b) Illustration of two metabolic feature pairs detected for the same metabolite. (a) Two mass spectra derived from positive and negative ionization mode for the respective native and corresponding $^{13}C_8$-labeled features derived from phenylalanine. (b) EIC profiles of the respective metabolic features shown in part a. (c) m/z versus retention time plot of all $^{13}C_9$-Phe-derived features detected in the positive and negative ionization mode and (d) their convolution into a feature group. The red dots represent selected metabolic features from three of the either annotated or identified metabolites. For details, see Supporting Information S1.2.

graphic peak shapes but can be easily distinguished by MS because of their differing m/z values. The observed m/z shift between the native and the partly tracer-labeled biotransformation product is proportional to the number of atoms of the labeling isotope in the remaining part of the tracer of the inspected metabolite ion. The presented workflow automatically searches for these unique isotope patterns and returns a comprehensive list of metabolic features, each corresponding to an ion of a metabolite derived from the studied tracer. As long as the isotope patterns of the tracer(s) incorporated in the respective metabolite can be separated, the algorithm can detect all metabolic feature pairs of a biotransformation product, including those isotopologs originating from the incorporation of different tracer moieties. The SIL-derived isotope patterns provide a high certainty that the detected metabolites are truly derived from the studied tracer.

To demonstrate the workflow, the metabolic fate of the amino acid phenylalanine (Phe) was studied in *Tae* cell suspensions cultured in the presence of U-$^{13}C_9$ Phe in the culture medium. Processing of the acquired raw data resulted in a total of 341 metabolic features, which were convoluted to 139 feature groups, each of which is representing a metabolite. Figure 2a shows two mass spectra containing a metabolism product with eight tracer-derived carbon atoms (native M) and its partly ^{13}C-labeled analog M′. The presence of the M′ + 1 mass peaks indicate that the moiety conjugated to the tracer also contains several carbon atoms.

Two medium blanks and *Tae* cell suspension cultures (no U-$^{13}C_9$ Phe added) were processed as described above. Only 1−2 of incorrectly detected metabolic features were found per sample, which during manual curation showed to be Fourier transformation artifacts with low abundances and noisy chromatographic peak shapes. This very low number of false positives confirms the high selectivity of the presented SIL assisted approach.

Fast Polarity Switching. The presented workflow supports fully automated processing of LC−ESI-HRMS data employing fast polarity switching. The cycle time for two successive MS

dx.doi.org/10.1021/ac503290j | *Anal. Chem.* 2014, 86, 11533−11537

scans (positive and negative ionization mode) was ~1 s, which is sufficient to acquire 10−25 scans per chromatographic peak. As shown in Figure 2b, the chromatographic peak shapes are very similar for all four depicted mass traces (of M_p, M_p', M_n, M_n'); thus, even features originating from the same metabolite but recorded in different ionization modes can be convoluted automatically into a single feature group.

In the presented experiment, 150 and 191 metabolic features were detected in the positive and negative ionization modes, respectively (Figure 2c). Furthermore, feature grouping across the two ion polarity modes was carried out successfully and resulted in a total of 139 distinct feature groups (i.e., metabolites). Of those, 32 were exclusively found in the positive mode, and 58 metabolites were detected in the negative ionization mode only. In addition, 49 metabolites exhibited at least one metabolic feature in each of the both ionization modes (Figure 2d). These findings underline the significant benefit that has been gained with respect to metabolite coverage by integrating positive and negative mode data.

Metabolite Annotation. The presented workflow results in a list of metabolic features, each corresponding to a certain ion species (e.g., adduct, deprotonated molecule, or in-source fragment) of a metabolite. For 19 of the 49 detected phenylalanine-derived metabolites with complementary adducts from both ionization modes, annotation of their intact neutral molecule and, thus, the corresponding molecular weight was achieved only by integration of ion species related information from the respective opposite ionization mode. This further demonstrates the power of fast polarity switching for complementary metabolite annotation.

The database search revealed that 50 of the detected metabolites, several with the same molecular mass, were putatively annotated. Although metabolite annotation was not always unambiguous, the metabolites could be assigned to phenylpropanoids (n = 14), phenylpropanoid amides (n = 9), and flavonoids (n = 10), which are partly known to have antagonistic effects against *Fusarium* infection (Supporting Information S1.2). Detailed metabolite annotation and biological interpretation of results will be published elsewhere.

DON Treatment. After data processing with the developed workflow, hierarchical clustering analysis (HCA) showed two distinct clusters for the two conditions "control" and "treatment". Furthermore, the heatmap illustration indicates that the abundance of many metabolites is either increased or decreased in the DON-treated samples. This confirms that the DON treatment had a severe impact on the cells metabolic composition (Supporting Information S1.3).

CONCLUSION

In recent years, SIL has been increasingly used in many fields of targeted and untargeted metabolomics research. The presented SIL-assisted LC−HRMS-based workflow is well suited to study the metabolic fate of both endogenous and exogenous tracer substances. All metabolites derived from the studied native and ^{13}C-labeled tracers show unique isotope patterns, which enable their untargeted detection and provide high confidence that the detected metabolites are truly derived from the studied tracer substance. The presented approach is particularly suited for the investigation of secondary metabolism and can be applied to virtually any biological system without the need for extra equipment other than the labeled tracer compound(s). As demonstrated for wheat cell suspension cultures, the use of ESI in combination with fast polarity switching, the study of endogenous plant secondary metabolite precursors (for example, phenylalanine) directly results in a large number of complementary Phe-related metabolic features, which can be assigned to various structure classes. In conclusion, our data demonstrate the great potential of SIL-assisted workflows for the comprehensive and highly selective untargeted screening and annotation of metabolites truly derived from the studied tracer. The workflow supports valuable applications across many different fields of metabolomics research, such as drug, pesticide, toxin, or any other secondary metabolite precursor-related conversion. A software tool and technical assistance enabling the fully automated processing according to the presented strategy are available from the corresponding author. Moreover, this software tool will be published and made freely available as part of an even more comprehensive software package for the evaluation of SIL-derived metabolomics data in the near future.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Author

*E-mail: rainer.schuhmacher@boku.ac.at.

Author Contributions

¶B.K. and C.B. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Denise Schöfbeck, Sylvia Lehner, and Benedikt Warth for useful discussions. The Austrian Science Fund (SFB-Fusarium-3706-B11;3702) and the Austrian Ministry of Science and Research (OMICS Center Graz) are acknowledged for financial support.

REFERENCES

(1) Levsen, K.; et al. *J. Chromatogr., A* **2005**, *1067* (1−2), 55−72.
(2) Sandermann, H. *Pest Manage. Sci.* **2004**, *60* (7), 613−623.
(3) Zhang, H.; et al. *J. Mass Spectrom.* **2008**, *43* (9), 1191−200.
(4) Baillie, T. A. *Pharmacol. Rev.* **1981**, *33* (2), 81−132.
(5) Iglesias, J.; Sleno, L.; Volmer, D. A. *Curr. Drug Metab.* **2012**, *13* (9), 1213−1225.
(6) Klein, S.; Heinzle, E. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **2012**, *4* (3), 261−272.
(7) Cabaret, O.; et al. *Rapid Commun. Mass Spectrom.* **2011**, *25* (19), 2704−10.
(8) Li, F.; et al. *J. Proteome Res.* **2013**, *12* (3), 1369−1376.
(9) Hiller, K.; et al. *Bioinformatics* **2013**, *29* (9), 1226−1228.
(10) Chokkathukalam, A.; et al. *Bioinformatics* **2013**, *29* (2), 281−283.
(11) Huang, X.; et al. *Anal. Chem.* **2014**, *86* (3), 1632−1639.
(12) Bueschl, C.; et al. *Metabolomics* **2014**, *10*, 754−769.
(13) Kluger, B.; et al. *Anal. Bioanal. Chem.* **2013**, *405* (15), 5031−6.
(14) Chassy, A. W.; et al. *Food Chem.* **2015**, *166* (0), 448−455.
(15) Gunnaiah, R.; Kushalappa, A. C. *Plant Physiol. Biochem.* **2014**, *83C*, 40−50.
(16) Du, P.; Kibbe, W. A.; Lin, S. M. *Bioinformatics* **2006**, *22* (17), 2059−2065.
(17) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, *8* (1), 105.