

# CPEXtract, a Software Tool for the Automated Tracer-Based Pathway Specific Screening of Secondary Metabolites in LC-HRMS Data

Bernhard Seidl, Rainer Schuhmacher,\* and Christoph Bueschl

Cite This: *Anal. Chem.* 2022, 94, 3543–3552

Read Online

ACCESS |



Metrics &amp; More

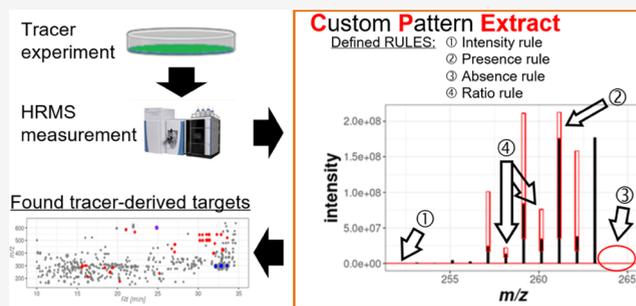


Article Recommendations



Supporting Information

**ABSTRACT:** The use of stable isotopically labeled tracers is a long-proven way of specifically detecting and tracking derived metabolites through a metabolic network of interest. While the recently developed stable isotope-assisted methods and associated, supporting data analysis tools have greatly improved untargeted metabolomics approaches, no software tool is currently available that allows us to automatically and flexibly search liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) chromatograms for user-definable isotopolog patterns expected for the metabolism of labeled tracer substances. Here, we present Custom Pattern Extract (CPEXtract), a versatile software tool that allows for the first time the high-throughput search for user-defined isotopolog patterns in LC-HRMS data. The patterns can be specified via a set of rules including the presence or absence of certain isotopologs, their relative intensity ratios as well as chromatographic coelution. Each isotopolog pattern satisfying the respective rules is verified on an MS scan level and also in the chromatographic domain. The CPEXtract algorithm allows the use of both labeled tracer compounds in nonlabeled biological samples as well as a reversed tracer approach, employing nonlabeled tracer compounds along with globally labeled biological samples. In a proof-of-concept study, we searched for metabolites specifically arising from the malonate pathway of the filamentous fungi *Fusarium graminearum* and *Trichoderma reesei*. 1,2,3-<sup>13</sup>C<sub>3</sub>-malonic acid diethyl ester and native malonic acid monomethyl ester were used as tracers. We were able to reliably detect expected fatty acids and known polyketides. In addition, up to 46 and 270 further, unknown metabolites presumably including novel polyketides were detected in the *F. graminearum* and *T. reesei* culture samples, respectively, all of which exhibited the user-predicted isotopolog patterns originating from the malonate tracer incorporation. The software can be used for every conceivable tracer approach. Furthermore, the rule sets can be easily adapted or extended if necessary. CPEXtract is available free of charge for noncommercial use at <https://metabolomics-ifa.boku.ac.at/CPEXtract>.



## 1. INTRODUCTION

Untargeted metabolomics aims at the unbiased detection of all metabolites produced by a biological system that is covered by the respective analytical method used. Typically both qualitative and quantitative differences in metabolites resulting from natural fluctuations, genetic modifications (e.g., wild-type (WT) versus knock-out), or experimental biotic or abiotic stress factors (e.g., untreated versus treated in any form) are then examined.<sup>1</sup> Nowadays, besides nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) is the most commonly used analytical technique to cope with this task. LC-HRMS allows the detection of a large number of different metabolites in the samples. However, usually, only a few of these can be identified by comparison with reference standards.<sup>2</sup> Moreover, it is often not even possible to reliably differentiate between signals derived from contaminants or artifacts and those of truly biological metabolites.<sup>3</sup> To encounter this problem, stable isotopically labeled samples, biological material, or tracer molecules can be employed. The

ingenious basic principle here is that labeling is incognito for the organism, i.e., labeled substances are metabolized almost the same way as unlabeled substances in biochemical reactions, but the label can be easily traced in analytical measurements.<sup>4,5</sup> Stable isotope-assisted (SIA) workflows were used as early as the 1950s in comprehensive studies of metabolic pathways, where the progression of biochemical signaling pathways and metabolic intermediates were investigated by tracking changes in the isotopic composition of metabolites of the metabolic pathway under study over time. This made it possible to study metabolic pathways and networks or to investigate the metabolic fate of substances (e.g., drugs or toxins) into

Received: October 19, 2021

Accepted: January 30, 2022

Published: February 15, 2022



intermediate or end products. Typically,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^2\text{H}$  isotopes are used for tracking derived metabolites through a given metabolic network of interest.<sup>5–9</sup>

Additionally, the labeling information can also be used for improved unknown compound annotation, e.g., by means of interpretation of SIA tandem mass spectrometry data.<sup>10</sup>

Meanwhile, software-supported approaches for the untargeted search for tracer-derived compounds are also commonly used in metabolomics, where in contrast to the more specific biosynthetic pathway studies, a snapshot of the entire metabolic state that can be captured by the method in use is analyzed.<sup>1</sup> In the latter case, the artificial isotope patterns caused by the metabolism of the labeled tracer are used to reliably track all tracer-derived metabolites that contain the entire tracer or parts thereof.

However, especially in untargeted SIA metabolomics studies, the huge amounts of raw LC-HRMS data do not allow the comprehensive, manual evaluation anymore. Therefore, a variety of (semi)automated SIA methods and data evaluation tools have recently been developed which greatly help to improve untargeted metabolomics approaches. Examples of such tools are MetExtract II,<sup>11</sup> X13CMS,<sup>12</sup> HiTIME,<sup>13</sup> geoRge,<sup>14</sup> ALLocator,<sup>15</sup> NTFD,<sup>16</sup> or Compound Discoverer 3X (Thermo Scientific). These software tools aim at the unbiased detection of all labeled metabolites and require a high degree of isotopic enrichment resulting in completely separated isotopolog patterns of the unlabeled and labeled metabolites, respectively. Each of these tools has been developed for particular, rigidly predefined isotope patterns or is based on direct comparison of native and labeling-derived isotope signatures and is therefore tailored to its respective specific application. For example, MetExtract II<sup>11</sup> and HiTIME<sup>13</sup> both require LC-HRMS data sets consisting of native and uniformly  $^{13}\text{C}$ -labeled compounds and thus detect either such labeled metabolites or their biotransformation products (e.g., mycotoxins, endogenous metabolites, or drugs). On the other hand, X13CMS,<sup>12</sup> geoRge,<sup>14</sup> or NTFD<sup>16</sup> aim at the untargeted search for the totality of all existing isotope-enriched metabolites that potentially have alterations in their isotope patterns after treatment with an isotopically labeled precursor (such as  $\text{U-}^{13}\text{C}_6$  glucose) in comparison to the same samples treated with the identical but nonlabeled precursor. These tools require the separate measurement of unlabeled and labeled samples. The assignment of labeled constituents is then based on the differential comparison of isotopolog patterns between labeled and unlabeled samples, and thus indirectly relies on correction for natural isotopic abundance to find all labeled compounds in the biological sample under investigation.

To the best of the authors' knowledge, however, there is no software tool available to date that allows the user to freely and flexibly specify a custom isotopolog pattern, which is then searched for in the LC-HRMS data set. Here, we present Custom Pattern Extract (CPEXtract) that allows the user to specify custom isotopolog patterns together with preset relative intensity ranges to be searched for, thereby enabling specific filtering for metabolite ions of interest that agree with the desired and diagnostic isotopolog pattern.

In contrast to the aforementioned software, CPEXtract does not have any rigid specifications but allows the user to define the isotopolog patterns to be searched for and the corresponding criteria freely via various rules. This allows, for example, the highly targeted search for a very specific

isotopolog pattern that can typically be expected to occur after multiple, iterative tracer incorporation during the biosynthesis of certain classes of specialized metabolites, as in the application presented here.

## 2. METHODS

**2.1. CPEXtract Application Example.** The application of CPEXtract is presented by searching specifically for metabolites of the malonate pathway, namely, fatty acids, polyketides, and possible hybrid metabolites thereof (e.g., nonribosomal peptide-polyketides (NRP-PKs) or prenylated polyketides) in culture samples of the filamentous fungi *Fusarium graminearum* and *Trichoderma reesei*.

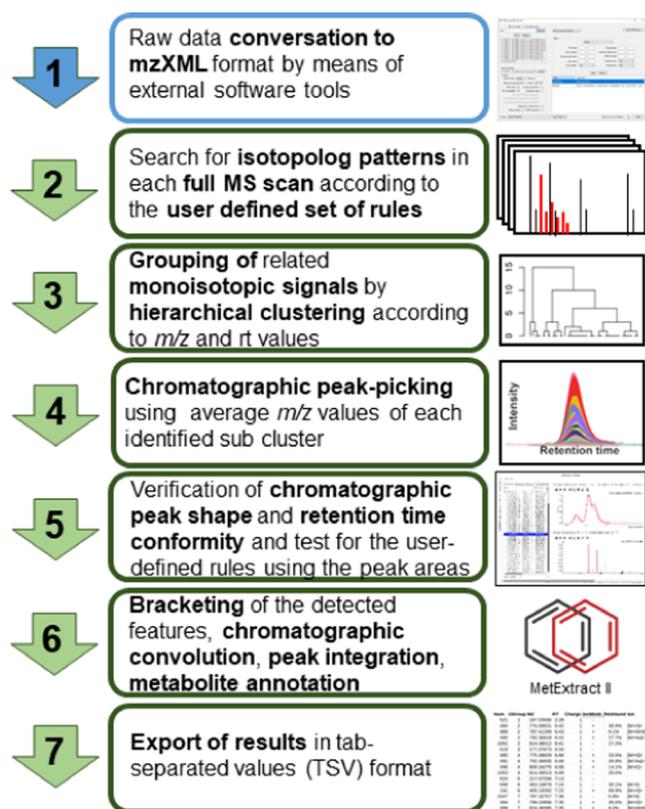
Two complementary tracer approaches were used. First, 1,2,3- $^{13}\text{C}_3$ -malonic acid diethyl ester was used as a tracer substance. This approach will henceforth be referred to as the standard tracer approach since the tracer was isotopically labeled ( $^{13}\text{C}$ ) and the cultivation was carried out with native glucose as the sole carbon source in the growth medium.

The second tracer substance was malonic acid monomethyl ester. Since this substance is not readily available in a  $^{13}\text{C}$  isotopically labeled form, a complementary approach was chosen, which will henceforth be referred to as the reversed tracer approach. Here, the native form of malonic acid monomethyl ester was used as a tracer, while the only other available carbon source for the fungus was  $\text{U-}^{13}\text{C}_6$ -labeled glucose. Detailed information about the cultivation workflow can be found in the [Supporting Information](#).

**2.2. CPEXtract Algorithm.** The automated CPEXtract workflow is based on MetExtract II software, which originally has been designed to detect pairs of native and uniformly  $^{13}\text{C}$ -labeled metabolite ions. CPEXtract replaces this fixed and rigid definition with a set of freely definable rules by which the isotope patterns to be searched for can be arbitrarily defined by the user. The individual rule set is defined via a Python vector of objects where each is an instance of one of the currently available "Rule" objects, which include the presence or absence of certain isotopologs as well as abundance or area-ratios between different isotopologs. Further custom user-defined rules can be easily added by defining additional subclasses. Automatic data processing then filters for all signals with isotope patterns that correspond to the defined set of rules. A schematic representation of the data processing workflow and algorithm is shown in [Figure 1](#).

The individual steps of the processing workflow and CPEXtract algorithms are:

1. To be processed, the manufacturer-specific raw data chromatograms must be converted into the mzXML format with, for e.g., ProteoWizard MSConvert.<sup>17</sup>
2. Data processing in CPEXtract starts at the MS scan level, whereby each individual  $m/z$  signal in a full scan mass spectrum is initially considered to represent the principal isotopolog of a putative metabolite ion of interest (henceforth termed X). This assumption is then verified by checking the user-defined set of rules. Currently, the following distinct rules are available: the "PresenceRule" (a particular isotopolog must be present and be within expected ratios relative to other isotopologs) and "AbsenceRule" (a specified isotopolog must not be present). For these two rules, a "minIntensity" parameter, which can be used to define a minimum signal intensity that the isotopolog must have, and a



**Figure 1.** Schematic overview chart of the data processing workflow and CPExtract algorithm.

“verifyChromPeaksSimilarity” parameter, which specifies the chromatographic peaks of the isotopologs, must be present and coelute. Furthermore, the “RatioRule” (checks whether an intensity ratio of two isotopologs or even two groups of several summed up isotopolog intensities corresponds to a certain defined ratio or checks if both ratios are within a defined range), the “AnyIntensityRule” and the “AllIntensityRule” (either any or all specified isotopologs must exceed a certain intensity threshold) are available. Moreover, users can easily implement their own rules if required. It is then checked if the isotopologs defined in the rule set relative to the assumed X isotopolog are valid (e.g., if certain isotopologs are present or absent). Only if X and all its putatively related isotopolog peaks meet the specified rules, X is used for subsequent data processing.

- All MS signals (i.e., Xs) that fulfilled the rules in the previous step are then clustered using hierarchical clustering. Subclusters of the thereby generated dendrogram within a low mass-deviation window (scan-to-scan variability) are kept. This dendrogram cutting step also takes the acquisition time (scan number according to the retention time) of the MS scans into consideration so that noncoeluting isomers end up in different sub-clusters.
- Data processing then continues on the chromatographic level. Therefore, an average  $m/z$  value is calculated for X of each subcluster present from the previous step, and peak picking is carried out using the continuous wavelet transform (CWT)-based centWave algorithm.<sup>18</sup>
- Only those isotope pattern candidates (or Xs) identified in the previous steps, whose chromatographic retention

and peak shape show sufficient agreement with the user’s rules, are ultimately accepted as hits. For this, all chromatographic peaks of the isotopologs must show perfect chromatographic coelution, which is tested via the Pearson correlation coefficient. Additionally also the isotopolog ratios based on the chromatographic peak areas are again checked and must meet the user’s rule criteria.

- The remaining steps of CPExtract are the same as those implemented in MetExtract II,<sup>11</sup> namely, bracketing of the detected features across all samples, chromatographic convolution (i.e., grouping and annotation of metabolite ions from the same metabolite), and peak area reintegration.
- Results of the automated data processing are provided in the form of a tab-separated value (TSV) file containing a data matrix with all detected principal isotopologs (Xs). The results can also be visualized and checked in the software’s graphical user interface, whereby both the extracted ion chromatograms and the spectra are illustrated.

**2.3. Fungal Strains and Cultivation Conditions.** *F. graminearum* PH-I (wild-type) obtained from Gerhard Adam, BOKU, Institute for Microbial Genetics (IMiG), and three different *T. reesei* strains, namely, QM6a $\Delta$ *tmus53* (wild-type),<sup>19</sup> QM6a  $\Delta$ *tmus53* $\Delta$ *sor1* (a knock-out mutant lacking the gene encoding the SOR1 polyketide synthase, the first enzyme in the sorbicillinoid biosynthetic pathway),<sup>20</sup> and QM6a  $\Delta$ *tmus53**ReYpr1* (a yellow pigment regulator 1 transcription factor overexpressing strain)<sup>21</sup> were used for the preparation of the culture medium supernatant and mycelium samples. To obtain the data sets, the samples were measured by LC-HRMS. Detailed information about the SLA-tracer workflow, cultivation, sample preparation, and measurement are given in the [Supporting Information](#).

**2.4. Conversion of LC-HRMS Raw Data Files.** The raw data files were converted into the mzXML format using the ProteoWizard MSConvert Software<sup>17</sup> version 3.0.19210 (settings: mzXML output format, 32-bit binary encoding precision, enabled index writing, enabled TPP compatibility, disabled zlib and gzip compression, peak picking using vendor algorithm).

**2.5. CPExtract Rules and Parameter Settings for Data Processing.** **2.5.1. Processing of LC-HRMS Data Generated Using the Standard Tracer Approach.** The following set of rules was used to detect isotopolog patterns of fully <sup>12</sup>C metabolites that incorporate several <sup>13</sup>C<sub>2</sub> units derived from the <sup>13</sup>C-labeled malonic acid diethyl ester tracer as expected from the standard tracer approach:

X represents the native, monoisotopic form of the metabolite ions and consists only of <sup>12</sup>C isotopes for each carbon atom, and X<sub>+v</sub> indicates that v <sup>12</sup>C atoms are exchanged for <sup>13</sup>C atoms in that particular isotopolog relative to X.

- X must have a signal abundance of at least 1E5 counts. X<sub>+2</sub> and X<sub>+4</sub> must have a signal intensity of at least 5E4 counts. This is used to consider only signals with a sufficient signal intensity and, for e.g., mask out signal noise (part of PresenceRule).
- As X already represents the monoisotopic <sup>12</sup>C isotope form of the metabolite, the signals for putative <sup>12</sup>C<sub>-1</sub> and <sup>12</sup>C<sub>-2</sub> isotopologs, therefore, must not be present, otherwise, they would be false positives. To avoid

false-negative results, signals with an intensity value of a maximum of 5% of X are accepted as putative noise signal for both isotopologs without a rule violation (AbsenceRule).

- The metabolite ion containing one and two  $^{13}\text{C}$ -malonate tracer-derived extension units  $X_{+2}$  and  $X_{+4}$  must be present and their abundance has to be within the range of 10–300% of X and  $X_{+2}$ , respectively. (RatioRule)
- Moreover, the metabolite ions containing one or two partial isotopically labeled extension units (respectively, one  $^{13}\text{C}$  and one  $^{12}\text{C}$ )  $X_{+1}$  and  $X_{+3}$  must also be present and must have an abundance of 10–200% relative to X and  $X_{+2}$  as well as  $X_{+2}$  and  $X_{+4}$ , respectively (RatioRule).
- The ratio of the abundance of X totaled with the abundance of  $X_{+2}$  to the abundance of  $X_{+1}$  must be the same as of the abundance ratio of  $X_{+2}$  totaled with the abundance of  $X_{+4}$  to the abundance of  $X_{+3}$  with a maximum permissible relative deviation of these ratios of  $\pm 10\%$ , respectively (i.e.,  $(X + X_{+2})/X_{+1} \approx (X_{+2} + X_{+4})/X_{+3}$ ). This rule takes advantage of the natural isotope distribution and the isotopic impurity of the tracer substance to remove false positives (RatioRule).
- The isotopologs X,  $X_{+1}$ ,  $X_{+2}$ ,  $X_{+3}$ , and  $X_{+4}$  must be present as coeluting chromatographic peaks (Pearson correlation  $\geq 0.85$ ) (part of the first PresenceRule).

**2.5.2. Processing of LC-HRMS Data Generated Using the Reversed Tracer Approach.** To detect isotopolog patterns of fully  $^{13}\text{C}$ -labeled metabolites that incorporate one or more  $^{12}\text{C}_2$  units derived from the native malonic acid monomethyl ester, slightly modified rules compared to the standard tracer approach were used.

In the case of the reversed tracer approach, X represents the uniformly  $^{13}\text{C}$ -labeled form of the metabolite ions and consists only of  $^{13}\text{C}$  isotopes for each carbon atom and  $X_{-v}$  indicates that  $v$   $^{13}\text{C}$  atoms are exchanged for  $^{12}\text{C}$  atoms.

- X must have a signal abundance of at least 1E5 counts.  $X_{-2}$ ,  $X_{-4}$ , and  $X_{-6}$  must have a signal intensity of at least 5E4 counts (PresenceRule).
- Since here X already represents the fully labeled  $^{13}\text{C}$  isotope form of the metabolite, the signals for putative  $^{13}\text{C}_{+1}$  and  $^{13}\text{C}_{+2}$  isotopologs should not be present. Signals up to 5% of X were tolerated (AbsenceRule).
- The metabolite ion containing one, two, and three  $^{12}\text{C}$ -malonate tracer-derived extension units  $X_{-2}$ ,  $X_{-4}$ , and  $X_{-6}$  must be present and their abundance has to be in the range of 10–300% of X,  $X_{-2}$ , and  $X_{-4}$ , respectively (RatioRule).
- Moreover, the metabolite ions containing one or two extension units with one  $^{12}\text{C}$  and one  $^{13}\text{C}$  (i.e.,  $X_{-1}$ ,  $X_{-3}$ , and  $X_{-5}$ ) must be present and must have an abundance of 10 to 200% relative to X and  $X_{-2}$ ,  $X_{-2}$  and  $X_{-4}$ , and  $X_{-4}$  and  $X_{-6}$ , respectively (RatioRule).
- The ratio of the abundance of X totaled with the abundance of  $X_{-2}$  to the abundance of  $X_{-1}$  must be the same as of the abundance ratio of  $X_{-2}$  totaled with the abundance of  $X_{-4}$  to the abundance of  $X_{-3}$  as well as the abundance ratio of  $X_{-4}$  totaled with the abundance of  $X_{-6}$  to the abundance of  $X_{-5}$  with a maximum permissible deviation of these ratios of  $\pm 10\%$ , respectively (i.e.,  $(X + X_{-2})/X_{-1} \approx (X_{-2} + X_{-4})/X_{-3} \approx (X_{-4} + X_{-6})/X_{-5}$ ) (RatioRule).

- All isotopologs defined via “PresenceRule” instructions ( $X$ ,  $X_{-1}$ ,  $X_{-2}$ ,  $X_{-3}$ ,  $X_{-4}$ ,  $X_{-5}$ , and  $X_{-6}$ ) had to be present as coeluting chromatographic peaks and must show a highly similar peak shape (Pearson correlation  $\geq 0.85$ ) (part of the first PresenceRule).

**2.5.3. General Processing Parameters.** The further data processing parameter settings of CPEXtract were: intra-scan-mass-accuracy:  $\pm 5$  ppm, inter-scan-mass-deviation:  $\pm 8$  ppm, EIC-extraction window:  $\pm 5$  ppm, retention-time-window: 3–36 min, minimum/maximum chromatographic peak width: 5–25 s, minimum Pearson correlation for coeluting chromatographic peaks: 0.85, maximum allowed retention-time-deviation for bracketing of multiple measurements: 0.1 min, and maximum allowed  $m/z$  deviation for bracketing: 5 ppm.

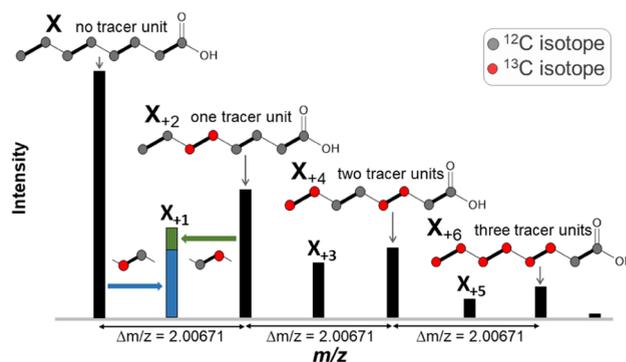
### 3. RESULTS AND DISCUSSION

A novel software tool, named CPEXtract, for the comprehensive and automated search for isotope patterns in LC-HRMS data that can be freely defined by the user is presented. Other than the existing software tools, it provides the user with an automated framework to mine the LC-HRMS data for certain characteristic isotopolog patterns via a set of predefined and extensible rules. Only isotopolog patterns obeying these rules are reported. Moreover, the large flexibility in defining the target patterns enables us to use the CPEXtract algorithm in two complementary ways, a standard tracer-, as well as a reversed tracer approach. As a proof-of-concept, CPEXtract software was used to search specifically for fungal metabolites of the malonate pathway, namely, fatty acids and polyketides as well as their derivatives like, for e.g., nonribosomal peptide-polyketide (NRP-PK) or prenyl-polyketide hybrids in LC-HRMS data.

#### 3.1. Expected Theoretical Isotope Patterns.

**3.1.1. Standard Tracer Approach.** Figure 2 shows a typical

##### Theoretical isotope pattern using the standard tracer approach



**Figure 2.** Theoretical isotope pattern of caprylic acid, which can be expected as a result of the incorporation of  $\text{C}_2$ -extension units derived from both, native malonate (synthesized by the fungus itself) and stable isotopically labeled malonate (tracer added to the medium) into the growing hydrocarbon chain.

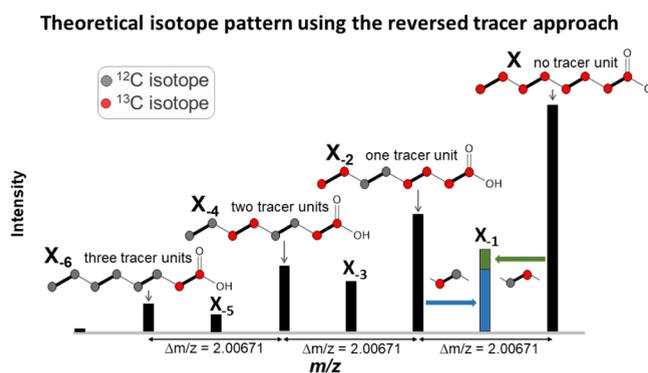
theoretical isotopolog pattern of a substance that is to be expected for the standard tracer approach, due to the accidental incorporation of both added ( $^{13}\text{C}$ -) tracer-derived and native fungus-produced  $\text{C}_2$  building blocks into its carbon skeleton. Such a pattern can be expected when the bioavailability and incorporation rate of tracer molecules is small in relation to the amount of malonyl-CoA units,

endogenously formed from native glucose and used by the fungus itself for biosynthesis.

A staircase-shaped isotope pattern is characteristic of continuously produced metabolites, corresponding to the decreasing probability that an increasing number of tracer-derived  $C_2$  units become incorporated into the synthesized metabolite.

In addition, also the  $X_{+1}$ ,  $X_{+3}$ , and  $X_{+5}$  isotopologs are of high diagnostic value. They result only from the proportion of non-monoisotopic tracer-derived extension units incorporated into the metabolites that occur due to the natural abundance of  $^{13}C$  isotopes in the native carbon source and the  $^{12}C$  isotopic impurity of the  $^{13}C$ -labeled tracer compound used. Therefore, their signal intensities must be smaller than the mean value of the two neighboring isotopologs with an even-numbered mass increment (e.g.,  $X_{+1} < (X + X_{+2})/2$ ). Furthermore, since the natural isotope abundance and isotopic purity of the tracer are constant, the resulting intensity ratio (e.g.,  $(X + X_{+2})/X_{+1}$ ) must be almost the same for all analogous isotopolog pairs (e.g.,  $(X_{+2} + X_{+4})/X_{+3}$ ). The intensity ratio of odd-numbered isotopologs to their even-numbered neighbors is independent of the number of tracer-derived extension units incorporated into the particular metabolite ion, making them ideal for highly specific filtering.

**3.1.2. Reversed Tracer Approach.** For experiments that require the use of compounds that are not available in the  $^{13}C$ -labeled form or would be too expensive, a reversed approach provides a solution. This approach combines global  $^{13}C$  labeling of the biological system with the use of a native tracer compound (such as malonic acid monomethyl ester). The thereby generated isotope patterns are a mirror image inversion of the patterns expected from the standard tracer approach (Figure 3).



**Figure 3.** Theoretical isotope pattern of caprylic acid, which can be expected as a result of the incorporation of  $C_2$  units derived from both,  $^{13}C$  malonate (synthesized by the fungus itself as grown on  $U\text{-}^{13}C_6$ -glucose as the sole carbon source) and native malonate (tracer added to the medium) into the growing hydrocarbon chain.

Due to the high flexibility in defining rules for the isotope patterns to be searched for, CPEXtract software works with both tracer approaches.

**3.1.3. Exemplification of Isotope Patterns and Tolerance Intervals for Oleic Acid.** Figures 4 and 5 show the spectra of oleic acid experimentally measured in *F. graminearum* mycelium extract samples together with the identified monoisotopic isotopolog ( $X$ ) and the respective tracer-derived isotopologs exemplarily for the standard tracer- and reversed tracer approach, respectively. For the *T. reesei* mycelium

extract, similar patterns were observed (data not shown). In both illustrations, the red boxes show the required isotopologs as well as their permissible intensity ranges as defined in the respective rule set for the standard- and the reversed tracer approach.

As can be seen in Figures 4 and 5, the rules selected are suitable for the detection of the isotope patterns in the measurement data caused by the random incorporation of tracer-derived  $C_2$  extension units. Due to the rapidly decreasing signal intensity of isotopologs with an increasing number of tracer-derived extension units, not all theoretically existing isotopologs were detected.

For this reason, the rule set of the standard tracer approach required the presence and permissible intensity of at least two  $C_2$  extension unit isotopologs ( $X_{+2}$  and  $X_{+4}$ ) as well as the isotopologs  $X_{+1}$  and  $X_{+3}$  in between for a hit. In addition, the intensity ratios of the sum of  $X$  plus  $X_{+2}$  to  $X_{+1}$  and  $X_{+2}$  plus  $X_{+4}$  to  $X_{+3}$ , respectively, had to be equal with a maximum permissible deviation of  $\pm 10\%$ .

With the reversed tracer approach, more abundant signals for the  $X_n$  isotopolog signals ( $X_{-1}$ ,  $X_{-2}$ , ...) were obtained compared to the standard tracer approach. Therefore, the rule set was extended by an additional pair of isotopologs corresponding to a third tracer-derived  $C_2$  extension unit ( $X_{-5}$  and  $X_{-6}$ ), further increasing the specificity.

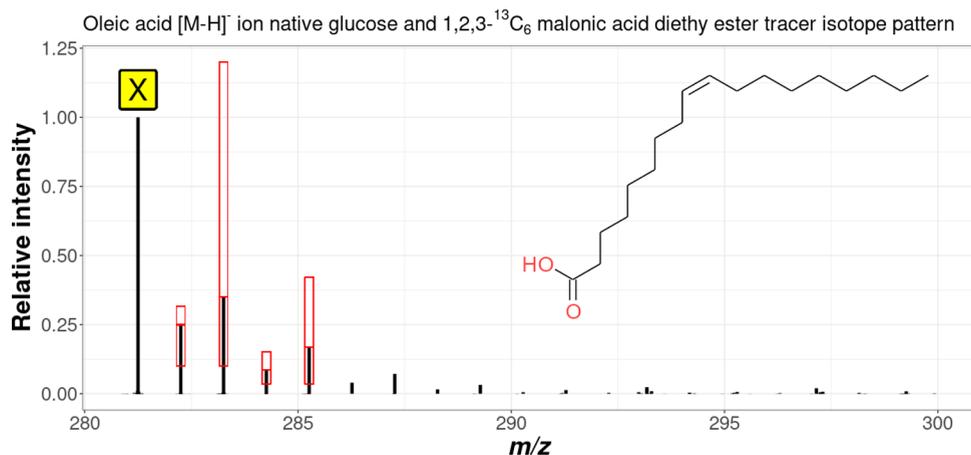
**3.2. Isotope Patterns Found for the Polyketides Aurofusarin and Sorbicillinol.** The experimentally observed isotopolog patterns after data processing with CPEXtract are exemplified with the two known polyketides aurofusarin for *F. graminearum* and sorbicillinol for *T. reesei*. The results are shown for both tracer approaches in Figure 6.

The spectrum of aurofusarin from the reversed tracer approach showed all 15 isotopolog signals corresponding to the incorporation of all 15 possible malonate tracer-derived  $C_2$  extension units into the carbon skeleton. In contrast, for aurofusarin in the standard tracer approach and for sorbicillinol in both approaches, only some of the theoretically possible isotopologs were sufficiently abundant for being detected.

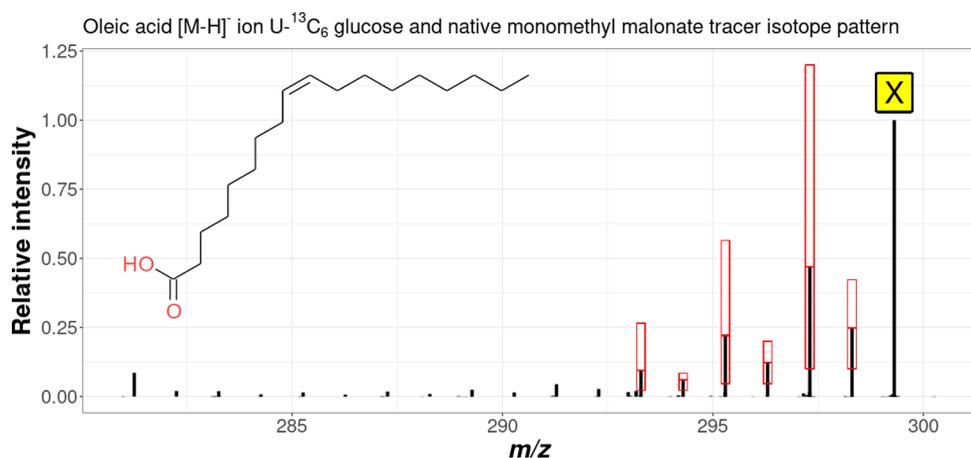
**3.3. Overall Pattern Shapes Found for the Polyketide Aurofusarin.** In both approaches, the intensity distribution does not necessarily have to follow the strict declining or increasing pattern shown above. Depending on the onset and time period of the metabolite synthesis during cultivation and the prevailing ratio of the added tracer to extension units formed by the fungus from the general carbon source, a different intensity profile can also appear. Moreover, the shape of the pattern also depends on whether a metabolite is possibly composed of two or more building blocks that were initially synthesized independently, as well as the cell compartment localization and time and rate of excretion of the biosynthesized product. Figure 7 illustrates, for e.g., the slightly differing isotope patterns obtained for the *F. graminearum*-derived polyketide aurofusarin ( $C_{30}H_{18}O_{12}$ ) in the mycelium extract and supernatant samples.

Consequently, the rules for the permissible intensity ranges for the  $X_{\pm 2}$  isotopologs must allow adequate leeway for being able to detect all metabolite ions of the target substance class.

While the large tolerance windows needed for the initial pattern search can potentially compromise the filtering specificity and thus lead to false-positive search results, additional pattern shape characteristics can be used for further filtering of candidate ions.



**Figure 4.** Measured spectrum of the  $[M - H]^-$  ion of oleic acid in the mycelium extract samples of *F. graminearum* using the standard tracer approach. X denotes the predominant, monoisotopic, native deprotonated molecule. Also shown in red boxes are the isotopologs defined in the rules that must be present and their permitted intensity ranges.



**Figure 5.** Measured spectrum of the  $[M - H]^-$  ion of oleic acid in the mycelium extract samples of *F. graminearum* using the reversed tracer approach. X denotes the predominant, monoisotopic fully  $^{13}\text{C}$ -labeled deprotonated molecule. Also shown in red boxes are the isotopologs defined in the rules that must be present and their permitted intensity ranges.

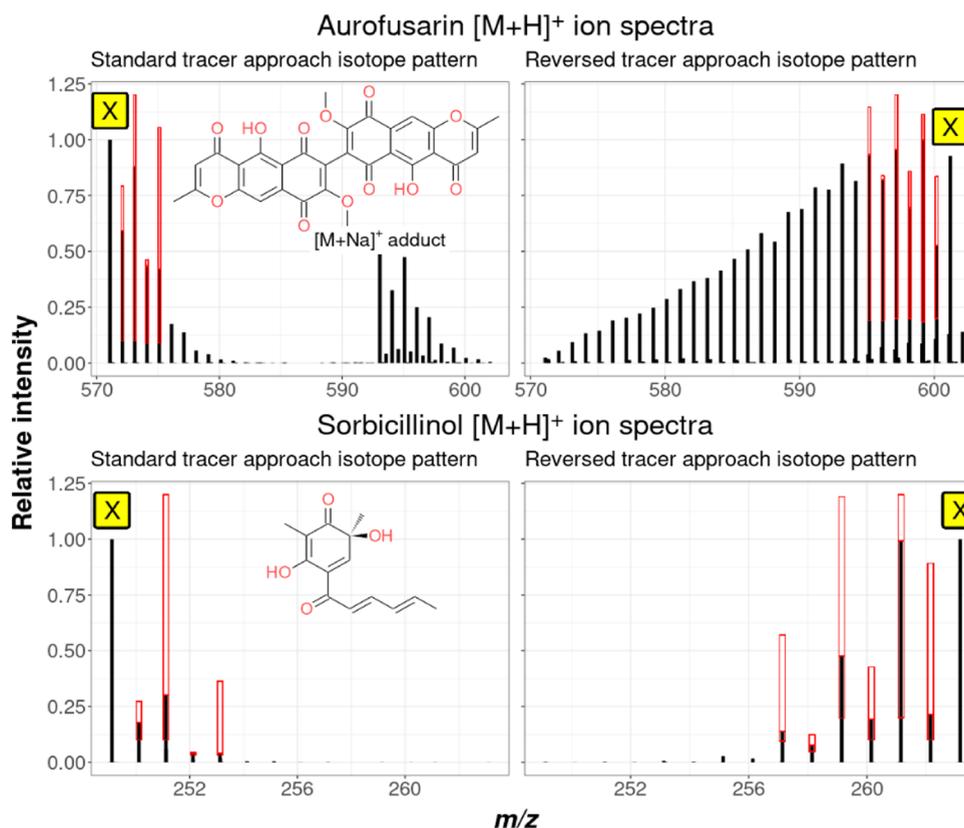
Insufficient predictability of the overall shape of the pattern does not affect the respective intensities of the isotopologs located between two adjacent even-numbered isotopologs (e.g., for  $X_{+1}$  between X and  $X_{+2}$ ). As already explained above, the odd-numbered ( $X_{+1}$ ,  $X_{+3}$ ,...) isotopologs always result from the sum of the respective isotope fractions and the number of carbon atoms in the metabolite. Furthermore, and perfectly usable as a reliable selection criterion, the relative ratio of the intensity of these isotopologs to that of the respective two analogous neighboring isotopologs with a differing number of tracer-derived extension units must remain the same. Therefore, much smaller tolerable intensity intervals can be defined in the rules for the odd-numbered isotopologs without running the risk to miss target metabolite ions. To this end, the initially created candidate list can be refined by checking the constant relative intensities as described above to sort out the remaining false positives very reliably.

**3.4. Global Screening of All Tracer-Derived Metabolite Ions.** The big advantage of filtering MS features according to certain metabolite classes by CPEXtract using specific isotope patterns derived from the corresponding tracer can clearly be seen in Figure 8. Likewise, the comparability of the two complementary tracer approaches is illustrated.

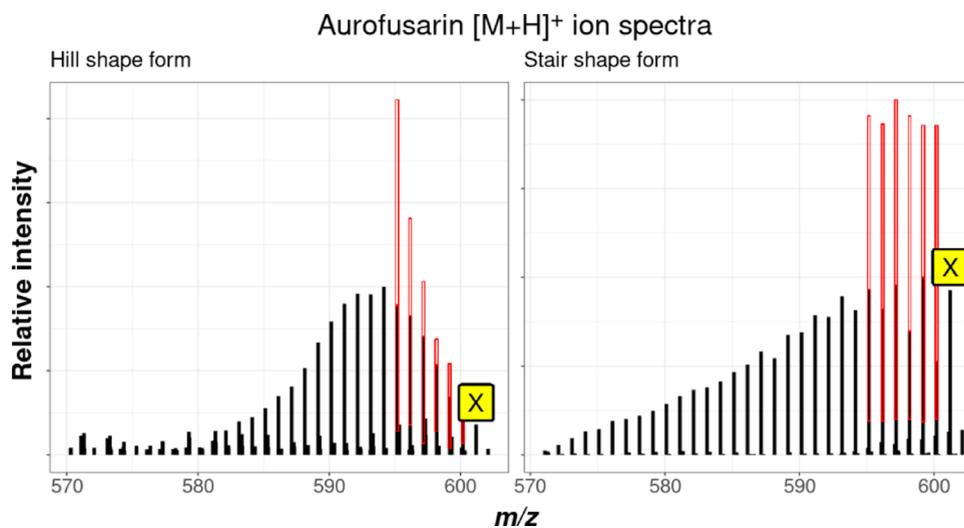
In the *F. graminearum* data set, the MetExtract II evaluation of the global labeling approach resulted in a total number of 621 metabolites found both in the supernatant and mycelium extract samples of the wild-type strain PH-I that were actually produced by the fungus. The CPEXtract algorithm detected a subset of tracer-derived metabolites from this set, comprising 39 metabolites for the standard tracer approach and 54 metabolites for the reversed tracer approach. In total, 46 of these metabolites could not be clearly identified with reference standards and thus represent possible novel metabolites.

This clearly demonstrates that the filter effect of the tracer approach combined with the automated CPEXtract data evaluation on the targeted metabolic pathway worked properly.

The importance of each rule in defining the isotopic pattern to be searched for became apparent as each rule was added step by step. When using the intensity rule only, 122 and 91 possible targets were detected in the data of the standard and reversed tracer approaches, respectively, including a large number of false positives. After adding the absence rules, many of these false positives were removed and 83 (standard approach) or 64 (reversed approach) possible targets remained. After adding the ratio rules, the remaining false positives were removed, leaving 54 and 39 found metabolites



**Figure 6.** Spectra of the metabolites aurofusarin (*F. graminearum*) and sorbicillinol (*T. reesei*), actually found with CPExtract for both tracer approaches. Also shown in red boxes are the isotopologs defined in the rules that had to be present and their permitted intensity ranges.



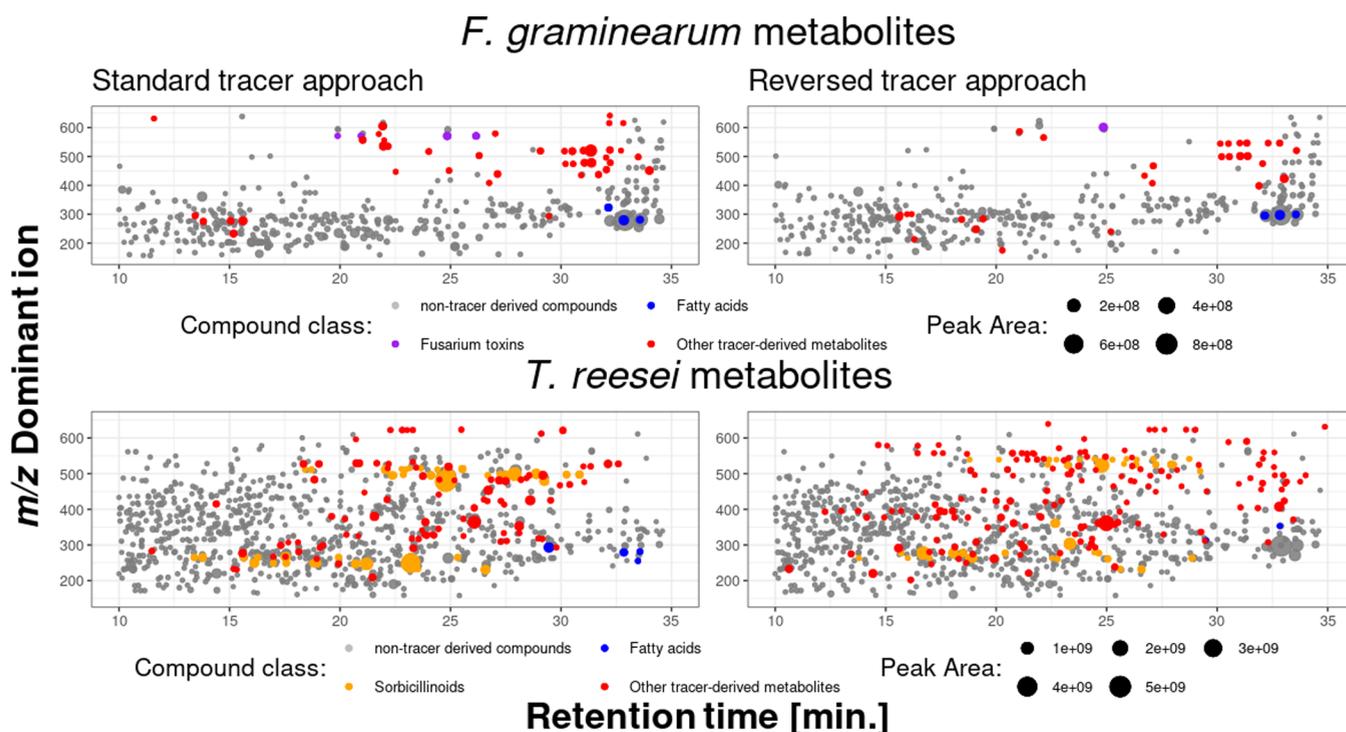
**Figure 7.** Different isotope pattern shapes of [M + H]<sup>+</sup> isotopologs of *F. graminearum*-derived aurofusarin. A hill-shaped form was found in mycelium extracts (left), whereas a stair-shaped form was found in supernatant samples (right).

in the standard and reversed tracer approach samples, respectively.

In the *T. reesei* data set, consisting of the supernatant and mycelium extract samples of each of the three different strains, 1165 metabolites truly originating from the fungus were found by MetExtract II data evaluation. The CPExtract algorithm data processing yielded 167 tracer-derived metabolites for the standard tracer approach and 283 for the reversed tracer approach, respectively. In total, 270 metabolites could not be clearly identified with reference standards and thus represent

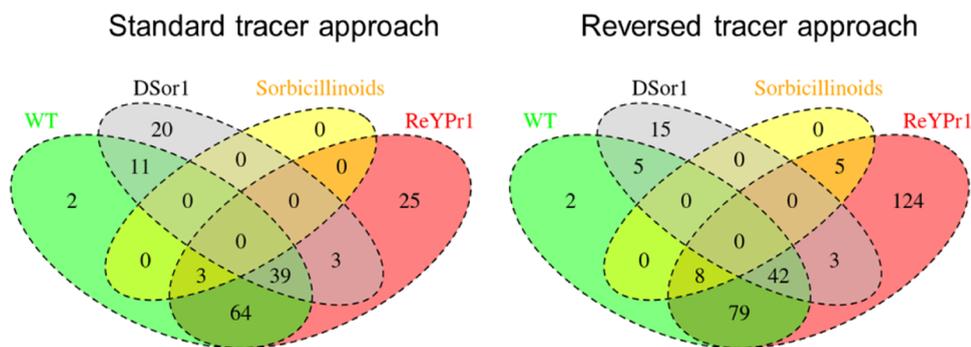
possible novel metabolites. When solely using the presence rules, 464 (standard approach) and 1462 (reversed approach) possible targets were detected. After adding the absence rules 282 (standard approach) or 418 (reversed approach) possible targets remained. The additional application of the ratio rule then yielded the remaining targets whose isotope pattern matched the expectation.

The relatively high number of (malonate-derived) metabolites in the *T. reesei* samples can be explained, on the one hand, by the combination of three different fungal genotypes and is



**Figure 8.** Mass-retention time plots of the dominant ion of each feature group from all truly fungal-derived metabolites found by MetExtract II software (shown in gray) in the sample data sets of *F. graminearum* and *T. reesei* (all three genotypes) culture samples. Metabolites detected by CPEXtract are shown in purple (known *F. graminearum* metabolites), yellow (known *T. reesei* metabolites), blue (fatty acids), or red (unknown, potential novel tracer-derived metabolites), illustrating the comparability of both tracer approaches and the filter effect of the targeted search for tracer-derived metabolites.

### Tracer-derived metabolites in the *T. reesei* culture samples



**Figure 9.** Venn diagrams of all metabolites found with both, the standard tracer and the reversed tracer approach in the investigated *T. reesei* strain culture samples (wild-type (WT) in green, Ypr1-overexpressing ReYpr1 strain in red, and  $\Delta$ Sor1 deletion strain in gray). The annotated sorbicillinoids are shown in yellow.

on the other hand presumably due to the high reactivity of the sorbicillinoids. Members of this class of polyketides are known to spontaneously react nonenzymatically into various other substances or reaction by-products. Figure 9 shows the all found feature groups detected in the *T. reesei* culture samples for both tracer approaches.

In the Venn diagrams, it can be seen that no metabolites annotated as sorbicillinoids were found in the  $\Delta$ Sor1 strain samples, which is as expected since this strain is deficient in the sorbicillinoid polyketide gene cluster and thus not able to produce any sorbicillinoids at all. Furthermore, additional sorbicillinoids (namely, putative bisorbicillinol, epoxyorbicillinol, sorbiquinol, bisvertinolone, dihydrobisvertinolone) were

found exclusively in samples of the overexpressing ReYpr1 strain.

**3.5. Selectivity of the CPEXtract Algorithm.** Furthermore, to test the selectivity of the CPEXtract algorithm and the rule sets used, culture samples without the addition of the tracer compounds were measured and processed using the same settings. This was done for both the standard- ( $^{12}\text{C}$  glucose) and reversed ( $^{13}\text{C}_6$ -glucose) approach. In these negative control samples, a low number of hits (false positives) were detected by CPEXtract, demonstrating the high specificity of the presented approach. Surprisingly, only in some *T. reesei* negative control samples, the corresponding tracer-derived isotope patterns were detected. In total, seven corresponding

tracer-derived patterns (different ones in each replicate) were detected in supernatant samples of the Ypr1-overexpressing ReYpr1 strain, and one tracer-derived pattern was detected in a single replicate of the WT supernatant samples. Manual inspection of the raw data files revealed that the false-positive signals were only detected immediately after the measurement of samples, which showed high signal intensities for exactly those ion species. Therefore, it was concluded that the false positives resulted from carryover into successive injections via the insufficiently rinsed injection needle of the HPLC system. Thus, the false-positive hits were not caused by a low selectivity of the CPEXtract algorithm but correctly detected and annotated as polyketides (data not shown).

#### 4. CONCLUDING REMARKS

Based on the exemplary search for metabolites of the malonate pathway in the culture samples of two different filamentous fungi using two complementary tracer approaches, it was possible to demonstrate the reliable function of the CPEXtract algorithm and its universal and versatile applicability. Tracer-derived metabolites were detected by searching for very specific isotope patterns that do not occur naturally as a result of the random incorporation of tracer molecules during biosynthesis. The pattern to be searched for, was defined by means of rules describing it, which CPEXtract then used for the search. Through the detection of unknown metabolites, specifically from the substance group of interest, in addition to the known and expected metabolites of the respective fungus, it is evident that CPEXtract software can also benefit the search of various other putatively novel or even completely unknown metabolites. Especially the timely but still demanding topic of natural product discovery is a promising field of application of CPEXtract. As exemplified in this article, the reliable filtering of the MS data for the fatty acids and polyketides searched for, enables a drastic reduction in the complexity of the global metabolite profiles. This can help to greatly simplify manual data curation that is subsequently required.

CPEXtract also shows great potential for other natural product classes such as nonribosomal peptides or terpenoids when being combined with other tracer compounds such as amino acids or mevalonate. Another interesting option is to screen for functionalization of natural products like, for example, methylations using L-methionine-(methyl-<sup>13</sup>C). Finally, the metabolic fate of (labeled) potentially toxic xenobiotics can be probed by CPEXtract either with the standard or reverse approach.

We anticipate that the presented tracer approach in combination with the flexible and easy-to-use CPEXtract software will be of interest to the metabolomics community working in related fields of research.

CPEXtract is available free of charge for noncommercial use at <https://metabolomics-ifa.boku.ac.at/CPEXtract>. The source code is available at <https://github.com/chrBoku/CPEXtract>. Likewise, the LC-HRMS data can be obtained from <https://metabolomics-ifa.boku.ac.at/CPEXtract/#datasets>.

#### ■ ASSOCIATED CONTENT

##### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c04530>.

One section with additional text and three figures describing the SLA-tracer workflow; cultivation; sample preparation and measurement in detail; and showing screenshots of CPEXtract software (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Author

**Rainer Schuhmacher** – University of Natural Resources and Life Sciences, Vienna, Department of Agrobiotechnology (IFA-Tulln), Institute of Bioanalytics and Agro-Metabolomics, 3430 Tulln, Austria; [orcid.org/0000-0002-7520-4943](https://orcid.org/0000-0002-7520-4943); Email: [rainer.schuhmacher@boku.ac.at](mailto:rainer.schuhmacher@boku.ac.at)

##### Authors

**Bernhard Seidl** – University of Natural Resources and Life Sciences, Vienna, Department of Agrobiotechnology (IFA-Tulln), Institute of Bioanalytics and Agro-Metabolomics, 3430 Tulln, Austria; [orcid.org/0000-0003-1022-5388](https://orcid.org/0000-0003-1022-5388)

**Christoph Bueschl** – University of Natural Resources and Life Sciences, Vienna, Department of Agrobiotechnology (IFA-Tulln), Institute of Bioanalytics and Agro-Metabolomics, 3430 Tulln, Austria; [orcid.org/0000-0003-1729-9785](https://orcid.org/0000-0003-1729-9785)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.1c04530>

##### Author Contributions

B.S. and R.S. designed the experiments; B.S. carried out the biological experiments and the LC-HRMS analysis; and B.S. and C.B. implemented the software tool and processed the data set. All authors contributed to the biological interpretation and the implementation of the manuscript.

##### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

The authors would like to thank the Austrian Science Fund (projects ReSMe P-26733, playNICE ZK-74) and the Provincial Government of Lower Austria (project OMICS 4.0) for funding this work. The presented study contributes in part to the Ph.D. thesis of B.S. Furthermore, the authors would also like to thank Robert Mach, Astrid Mach-Aigner, and Christian Derntl (TU-Wien) for providing the *T. reesei* strains and Gerhard Adam and Gerlinde Wiesenberger (University of Natural Resources and Life Sciences, Vienna) for providing the *F. graminearum* strain.

#### ■ REFERENCES

- (1) Schuhmacher, R.; Krska, R.; Weckwerth, W.; Goodacre, R. *Anal. Bioanal. Chem.* **2013**, *405*, 5003–5004.
- (2) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098.
- (3) Sauer schnig, C.; Doppler, M.; Bueschl, C.; Schuhmacher, R. *Metabolites* **2018**, *8*, No. 1.
- (4) Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R. *Anal. Bioanal. Chem.* **2013**, *405*, 27–33.
- (5) Rinkel, J.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2015**, *11*, 2493–2508.
- (6) Baillie, T. A. *Pharmacol. Rev.* **1981**, *33*, 81–132.
- (7) Chokkathukalam, A.; Kim, D.-H.; Barrett, M. P.; Breitling, R.; Creek, D. J. *Bioanalysis* **2014**, *6*, 511–524.
- (8) Zamboni, N.; Saghatelian, A.; Patti, G. J. *Mol. Cell* **2015**, *58*, 699–706.

- (9) Crown, S. B.; Long, C. P.; Antoniewicz, M. R. *Metab. Eng.* **2016**, *38*, 10–18.
- (10) Doppler, M.; Bueschl, C.; Kluger, B.; Koutnik, A.; Lemmens, M.; Buerstmayr, H.; Rechthaler, J.; Krska, R.; Adam, G.; Schuhmacher, R. *Front. Plant Sci.* **2019**, *10*, No. 1366.
- (11) Bueschl, C.; Kluger, B.; Neumann, N. K. N.; Doppler, M.; Maschietto, V.; Thallinger, G. G.; Meng-Reiterer, J.; Krska, R.; Schuhmacher, R. *Anal. Chem.* **2017**, *89*, 9518–9526.
- (12) Huang, X.; Chen, Y. J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J. *Anal. Chem.* **2014**, *86*, 1632–1639.
- (13) Leeming, M. G.; Isaac, A. P.; Zappia, L.; O'Hair, R. A. J.; Donald, W. A.; Pope, B. J. *SoftwareX* **2020**, *12*, No. 100559.
- (14) Capellades, J.; Navarro, M.; Samino, S.; Garcia-Ramirez, M.; Hernandez, C.; Simo, R.; Vinaixa, M.; Yanes, O. *Anal. Chem.* **2016**, *88*, 621–628.
- (15) Kessler, N.; Walter, F.; Persicke, M.; Albaum, S. P.; Kalinowski, J.; Goesmann, A.; Niehaus, K.; Nattkemper, T. W. *PLoS One* **2014**, *9*, No. e113909.
- (16) Weindl, D.; Wegner, A.; Hiller, K. *Methods Enzymol.* **2015**, *561*, 277–302.
- (17) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. *Nat. Biotechnol.* **2012**, *30*, No. 918.
- (18) Tautenhahn, R.; Böttcher, C.; Neumann, S. *BMC Bioinf.* **2008**, *9*, No. 504.
- (19) Steiger, M. G.; Vitikainen, M.; Uskonen, P.; Brunner, K.; Adam, G.; Pakula, T.; Penttilä, M.; Saloheimo, M.; Mach, R. L.; Mach-Aigner, A. R. *Appl. Environ. Microbiol.* **2011**, *77*, 114–121.
- (20) Derntl, C.; Guzmán-Chávez, F.; Mello-de-Sousa, T. M.; Busse, H.-J.; Driessen, A. J. M.; Mach, R. L.; Mach-Aigner, A. R. *Front. Biol.* **2017**, *8*, No. 2037.
- (21) Derntl, C.; Rassinger, A.; Srebotnik, E.; Mach, R. L.; Mach-Aigner, A. R. *Appl. Environ. Microbiol.* **2016**, *82*, 6247–6257.