

# Non-targeted tracer fate detection

Daniel Weindl, Andre Wegner, Karsten Hiller\*

*Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 7, Avenue des Hauts-Fourneaux, L-4362 Esch-Belval, Luxembourg*

---

## Abstract

Stable isotopes have been used to trace atoms through metabolism and quantify metabolic fluxes for several decades. Only recently non-targeted stable isotope labeling approaches have emerged as a powerful tool to gain biological insights into metabolism. However, the manual detection of isotopic enrichment for a non-targeted analysis is tedious and time consuming. To overcome this limitation, the non-targeted tracer fate detection (NTFD) algorithm for the automated metabolome-wide detection of isotopic enrichment has been developed. NTFD detects and quantifies isotopic enrichment in the form of mass isotopomer distributions (MIDs) in an automated manner, providing the means to trace functional groups, determine MIDs for metabolic flux analysis, or detect tracer-derived molecules in general. Here, we describe the algorithmic background of NTFD, discuss practical considerations for the freely available NTFD software package, and present potential applications of non-targeted stable isotope labeling analysis.

**Keywords:** NTFD, non-targeted, isotope labeling, stable isotopes, tracer fate,  $^{13}\text{C}$ , carbon-13

---

\*Corresponding author

*Email address:* karsten.hiller@uni.lu (Karsten Hiller)

---

## 1. Introduction

Since the recognition of cellular metabolism as a potential intervention point to treat complex diseases, stable isotope-assisted metabolomics techniques have emerged as a valuable tool to analyze intracellular fluxes [1]. Stable isotope labeled tracers have been successfully applied to study metabolic fluxes in cell culture and *in vivo* in animals and humans [2–7].

In the past such analyses have been rather targeted, mainly because isotopic enrichment from stable isotopes is not as readily detected as, for example, enrichment from radioactive isotopes. After a stable isotope labeling experiment, isotopic enrichment is typically determined by comparing the mass isotopomer distribution (MID) observed by mass spectrometric analysis with the MID to be expected from known natural isotope abundances [8]. However, to determine the theoretical MID, knowledge of the elemental composition of the compound of interest is required [9]. For this reason, usually only a predefined list of compounds is analyzed for isotopic enrichment.

It is self-evident that such a targeted approach is suited to test a well defined hypothesis, but fails to generate knowledge on unknown or unexpected parts of the metabolic network. Although many metabolic pathways are well described and known since many years, new reactions are continuously being identified, [10, 11] underlining the necessity for a global detection of isotopic enrichment. Recently, several tools for the non-targeted detection of isotopic enrichment in mass spectrometric data have been developed [12–18].

Alternatively, isotopic enrichment can be determined using isotope ratio mass spectrometry (IRMS) [19] with high precision in a non-targeted

manner. Unlike conventional mass spectrometric methods, IRMS cannot be used to determine MIDs. However, MIDs can provide interesting insights into metabolism and are the basis for  $^{13}\text{C}$ -metabolic flux analysis ( $^{13}\text{C}$ -MFA) [20, 21].

Here we will present the non-targeted tracer fate detection (NTFD) algorithm [12] which is able to detect and quantify isotopic enrichment in GC-EI-MS data in a non-targeted manner. We will introduce the theoretical background necessary to obtain meaningful results, discuss practical considerations, and present potential applications of non-targeted stable isotope labeling analysis.

< < < < <

NOTE TO TYPESETTER – PLACE THIS TEXT INTO A TEXT BOX

### **Isotopic isomers**

Stable isotope labeling leads to isotopic isomerism. Usually, the following classes of isotopic isomers are distinguished:

**Isotopologues** The widest class of isotopic isomers. Isomers that differ only in their isotopic composition with otherwise identical structure, including isotopic substitutions of different elements [22]. For example  $[\text{U-}^{13}\text{C}]\text{acetate}$  ( $[\text{1,2-}^{13}\text{C}_2]\text{acetate}$ ),  $[\text{1-}^{13}\text{C}]\text{acetate}$ , and  $[\text{2,2,2-D}_3]\text{acetate}$  are isotopologic. In carbon labeling experiments, the term is often implicitly used to only refer to  $^{13}\text{C}$ -isotopologues.

**Mass isotopomers** Mass isotopomers are groups of isotopologues of the same nominal mass or of isotopologues that cannot be separated by

mass spectrometry [23]. Mass isotopomers are referred to as  $M_i$ ,  $Mi$ , or  $M+i$ , where  $i$  indicates the increase in nominal mass as compared to the lightest isotopologue. For example, for carbon labeling of acetate,  $M_1$  comprises  $[1-^{13}\text{C}]\text{acetate}$  and  $[2-^{13}\text{C}]\text{acetate}$ .

**Isotopomers** Isotopomers are isotopologues with the same isotopic substituents but in varying positions [24]. For example,  $[1-^{13}\text{C}]\text{acetate}$  and  $[2-^{13}\text{C}]\text{acetate}$ , but not  $[1,2-^{13}\text{C}_2]\text{acetate}$  are isotopomeric. Isotopomers cannot be separated by mass spectrometry.

**MID** The mass isotopomer distribution (MID) is the relative abundance of the different mass isotopomers of a molecule. MIDs can be determined from mass spectrometric measurements and are often the basis for  $^{13}\text{C}$ -MFA, a method for metabolic flux analysis based on stable isotope labeling and stoichiometric modeling, where a set of fluxes through a predefined metabolic network is determined that can best explain the experimentally observed isotopic labeling.

> > > > >

## 2. Theoretical background

### 2.1. Overview

Before discussing experimental and analytical considerations for the use of NTFD, we will provide the theoretical background for the non-targeted detection and quantification of isotopic enrichment that is necessary for the

user to obtain meaningful results from NTFD. After the theoretical introduction we will explain which parameters of the NTFD implementation[13] are critical for each stage.

The NTFD algorithm operates on data from two MS measurements, one of an isotopically enriched (“labeled”) and one of a non-enriched (“unlabeled”) sample. After peak picking and mass spectral deconvolution, the mass spectrum of each compound from the labeled sample is paired with its counterpart from the unlabeled sample. Isotopically enriched compounds and mass spectrometric fragments are detected from the differences in these mass spectra and the isotopic enrichment is quantified (Figure 1). Along with the MID calculation, certain measures are determined to judge the quality of the obtained MID results.

The current NTFD software is designed to analyze low resolution GC-MS data. For high resolution LC-MS data, there are other tools available. These tools are beyond the scope of this chapter and we will only provide a short overview. Briefly, MetExtract [16] is a tool for the non-targeted detection of signals of unlabeled and uniformly labeled metabolites in LC-MS data. These information allow to differentiate between metabolites and analytical noise and the isotopologue ratios can be used for relative quantification of metabolite levels in two samples. However, MetExtract cannot be used to analyze complex mixtures of isotopologues. IsoMETLIN [17] is a web-based tool that identifies isotopologues of known compounds based on a computationally generated database and is thus rather targeted. It cannot identify unknown labeled metabolites. IsoMETLIN can make use of MS/MS spectra to resolve individual isotopomers. Such information can also be ob-

tained from NTFD by analyzing specific mass spectrometric fragments [9] or combinations thereof [25]. MzMatch-ISO [18] and X<sup>13</sup>CMS [14] are both R packages that allow for the non-targeted determination of MIDs and can be regarded as the LC-MS equivalents of NTFD.

We will now discuss the NTFD algorithm for the detection and quantification of isotopic enrichment in more detail.

## *2.2. Detection of isotopically enriched fragments*

When a compound is isotopically enriched, its mass spectrum changes. Assuming there are no isotope effects on ionization and fragmentation, the relative intensities of all fragments will be conserved. However, within each enriched isotopic peak cluster, the relative  $M_0$  intensity will decrease and the intensity of the remaining peaks will increase as compared to the non-enriched spectrum (Figure 2). The decrease in  $M_0$  equals the absolute value of the summed increase in heavier mass isotopomer abundances, since we assume that the relative fragment intensities remain constant.

These changes become more obvious in the difference spectrum where a labeled fragment will lead to a positive peak followed by a negative peak with the same area under the peak (Figure 2). These characteristic patterns are used to detect all labeled fragments.

For this detection of isotopic enrichment, the mass spectra of all compounds from the labeled and unlabeled sample need to be compared. Therefore, a spectrum and retention time (or retention index) matching [26] is performed between all mass spectra detected in the labeled and unlabeled samples during the previous step (Figure 1). To ensure only identical compounds are matched, this matching is highly stringent in terms of retention

time and loose in terms of mass spectrum similarity. Therefore, there should not be any significant shifts if retention times between the two measurements. The loose spectrum matching will ensure that the spectra of the labeled and unlabeled compound are also matched correctly in case of high isotopic enrichment, which can lead to very dissimilar mass spectra of the same compound.

In the case that replicate measurements have been provided, mass spectra from these chromatograms are matched first. This replicate mass spectra matching is highly stringent in terms of mass spectrum similarity, since there should not be any differences in isotopic enrichment and, thus, the mass spectra should be identical.

Once the mass spectra from the labeled ( $I_l^{raw}$ ) and unlabeled sample ( $I_{ul}^{raw}$ ) have been paired, they are analyzed for differences. Therefore, the peak intensities across the whole mass spectrum are normalized to a sum up to 1:

$$I = \frac{I^{raw}}{\sum_i I_i^{raw}} \quad (1)$$

The normalized spectra are then subtracted to yield the difference spectrum  $D$  (Figure 2):

$$D = I^{ul} - I^l \quad (2)$$

The detection of the characteristic pattern for the labeled fragment is more robust in the first derivative of the difference spectrum (Figure 2). Using a Savitzky-Golay filter with window size 5, a smoothed variant of the first derivative of the difference spectrum can be calculated very efficiently.

In this first derivative, isotopic enrichment will manifest as two positive peaks separated by a negative peak for each labeled fragment (Figure 2). These two subsequent maxima indicate the boundaries of a labeled fragment shifted by 2 mass units due to the previous smoothing step. Therefore, by scanning the first derivative for such patterns, all isotopically enriched fragments are determined along with their  $m/z$  ranges. These ranges are required for the next step to quantify the isotopic enrichment.

### *2.3. Quantification of isotopic enrichment*

The detection of isotopic enrichment in the previous step was only qualitative. This enrichment now needs to be quantified in terms of an MID for each labeled fragment. The raw MID observed in the mass spectrum of a compound after stable isotope labeling is a combination of natural isotope abundance and eventually from artificial isotopic enrichment as derived from the tracer (Figure 3).

When analyzing stable isotope labeling data, usually only the artificial enrichment derived from the tracer is of interest. Therefore, the raw mass isotopomer distribution as obtained from the mass spectrum needs to be corrected for natural isotope abundance. This is especially important in GC-MS analyses where often a chemical derivatization is part of the sample preparation, which introduces a large number of carbon or silicon atoms with high natural isotope abundances.

In targeted stable isotope labeling analyses this correction for natural isotope abundances is usually performed based on the theoretical MID as expected from the elemental composition and the average natural isotope abundance of each element [8]. However, this method is not applicable for

non-targeted applications, since many compounds and fragments cannot be identified and, thus, their elemental composition is not available.

Jennings and Matthews [27] presented an elegant method to correct raw MIDs for natural isotope abundance through comparison with the mass spectrum of the unlabeled compound. An adaptation of this approach is applied in NTFD.

In a  $^{13}\text{C}$ -labeling experiment, a compound with  $N$  carbons can contain maximally up to  $N$   $^{13}\text{C}$  isotopes from artificial tracer-derived enrichment. The remaining, non-enriched atoms are subject to natural isotope abundance (Figure 3A). The relative peak intensities in the mass spectrum of such a compound represent the raw MID. The raw MID  $I^l$ , of a fragment of such a labeled compound arises from natural MIDs of its “artificial” mass isotopomers as depicted in Figure 3B. The relative abundances in  $I^l$  are given as the summation of the natural MIDs of all  $N + 1$  artificial mass isotopomers  $M^0, \dots, M^N$ , weighted by their relative abundance or the corrected MID  $M$ :

$$I^l = \begin{pmatrix} M_0^0 \cdot M_0 \\ M_1^0 \cdot M_0 + M_0^1 \cdot M_1 \\ \vdots \\ M_n^0 \cdot M_0 + M_{n-1}^1 \cdot M_1 + \dots + M_0^n \cdot M_n \end{pmatrix} \quad (3)$$

$$= \begin{pmatrix} M_0^0 & 0 & \dots & 0 \\ M_1^0 & M_0^1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ M_n^0 & M_{n-1}^1 & \dots & M_0^n \end{pmatrix} \cdot M = \mathbf{M}_{corr} \cdot M \quad (4)$$

with  $M_b^a$  as the relative abundance of the  $b$ th natural mass isotopomer of

the compound containing  $a$  tracer isotopes. If the correction matrix  $\mathbf{M}_{corr}$  is known, the corrected MID can be determined as

$$M = \mathbf{M}_{corr}^{-1} \cdot I^l \quad (5)$$

To correct the raw MID for natural isotope abundance, we need to know the natural MIDs  $M^a$  of the artificial mass isotopomers. As mentioned above, they cannot be calculated from natural isotope abundances, since the elemental compositions is not available. However, we can obtain this information from the unlabeled mass spectrum  $I^{ul}$ . The first column of  $\mathbf{M}_{corr}$ ,  $M^0$ , is the natural MID of the molecule, without any artificial enrichment. These values are exactly the relative peak intensities in the mass spectrum  $I^{ul}$  of the unlabeled compound, leading us with (3) to:

$$\mathbf{M}_{corr} = \begin{pmatrix} s_0^{ul} & 0 & \cdots & 0 \\ s_1^{ul} & M_0^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_n^{ul} & M_{n-1}^1 & \cdots & M_0^n \end{pmatrix} \quad (6)$$

The remaining  $M_b^a$  can be replaced in a similar way. The mass spectra of  $M^1, \dots, M^N$  look similar to the unlabeled spectrum, but are shifted to higher mass by  $1, \dots, N$  (Figure 3). Furthermore, due to the fixed inclusion of stable isotopes in  $M^1, \dots, M^N$ , the  $M_0$  abundance will increase, and the other mass isotopomer abundances will decrease. Since the relative natural abundance of  $^{13}\text{C}$  and the number of carbons in most metabolites are rather low, we assume that only the  $M_0$  and  $M_1$  abundances are affected by the artificial enrichment, and neglect the insignificant changes in the abundances

of higher mass isotopomers.

When a carbon atom in a molecule is artificially substituted by a  $^{13}\text{C}$  isotope, this atom is no longer subject to natural isotope abundance, so that the  $M_0$  abundance of the molecule increases and the  $M_1$  abundance decreases. When  $a$   $^{13}\text{C}$  isotopes are included, the ratio of  $M_1$  and  $M_0$  changes to [27]:

$$\frac{M_1^a}{M_0^a} = \frac{M_1^0}{M_0^0} - a \cdot \frac{p_{^{13}\text{C}}}{p_{^{12}\text{C}}} \quad (7)$$

with  $p_{^{12}\text{C}}$  and  $p_{^{13}\text{C}}$  as the average natural isotope abundances of  $^{12}\text{C}$  and  $^{13}\text{C}$ , respectively. We define this correction term as:

$$c_a = a \cdot \frac{p_{^{13}\text{C}}}{p_{^{12}\text{C}}} \quad (8)$$

Since we assume that the relative abundances of the heavier mass isotopomers remain constant, the  $M_0$  abundance increases by the same value as the  $M_1$  abundance decreases. Therefore, their sum will not change:

$$M_1^a + M_0^a = M_1^0 + M_0^0 \quad (9)$$

Combining Equation 7, 8 and 9 provides a correction term  $M1_{corr}$  to obtain  $M_0^a$  from  $M_0^0$  as:

$$\frac{M_1^0}{M_0^0} \cdot M_0^a - c_a \cdot M_0^a = M_1^0 + M_0^0 - M_0^a \quad (10)$$

$$\frac{M_1^0}{M_0^0} \cdot M_0^a - c_a \cdot M_0^a + M_0^a = M_1^0 + M_0^0 + M_0^0 \cdot c_a - M_0^0 \cdot c_a \quad (11)$$

$$M_0^a = M_0^0 + \frac{M_0^0 \cdot c_a}{\frac{M_1^0}{M_0^0} + 1 - c_a} = M_0^0 + M1_{corr} \quad (12)$$

Since the sum of  $M_0$  and  $M_1$  need to be constant (Equation 9),  $M_1$  needs to decrease by the same amount as  $M_0$  increases:

$$M_1^a = M_1^0 - \frac{M_0^0 \cdot c_a}{\frac{M_1^0}{M_0^0} + 1 - c_a} \quad (13)$$

In summary, the natural mass isotopomer abundances  $M_b^a$  are given as:

$$M_b^a = \begin{cases} M_b^0, & \text{for } a = 0 \vee b > 1 \\ M_0^0 + M1_{corr}, & \text{for } a > 0 \wedge b = 0 \\ M_1^0 - M1_{corr}, & \text{for } a > 0 \wedge b = 1 \end{cases} \quad (14)$$

This enables us to populate the correction matrix  $\mathbf{M}_{corr}$  accordingly and to determine the isotopic enrichment from Equation 5 by least-squares regression without requiring any further knowledge on the given compound.

We explained this MID correction for the case of  $^{13}\text{C}$  labeling, but it can be applied to any other isotopic element. Depending of the natural isotope abundance of the isotopic element, a  $M_2$  correction may become necessary as described in Jennings and Matthews [27]. Combined labeling from different isotopic elements (*e.g.*  $^{13}\text{C}$  and  $^{15}\text{N}$ ) can not be corrected using this approach, since the contributions of the two isotopic species cannot be distinguished.

#### 2.4. Quality measures

To judge the quality of the calculated MIDs, NTFD provides three quality measures: the coefficient of determination, the sum of absolute values of the relative mass isotopomer abundances, and confidence intervals of the relative mass isotopomer abundances.

#### 2.4.1. Sum of absolute values

Theoretically, all relative mass isotopomer abundances should be within the interval  $[0, 1]$  and their sum should be equal to one:

$$\sum_i M_i = 1 \quad | \quad M_i \in [0, 1] \quad (15)$$

However, due to measurement errors, the regression can lead to negative mass isotopomer abundances. The sum of the absolute values

$$\sum_i |M_i| \quad (16)$$

can therefore be greater than one. This sum of absolute values provides a first, simple, yet effective measure for the accuracy of the calculated MID. The closer this value is to one, the more accurate are the MIDs.

#### 2.4.2. Coefficient of determination

If replicate measurements of the labeled and unlabeled sample are available, the coefficient of determination  $R^2$  can be computed for every fragment. This value is defined as

$$R^2 = \frac{\sum_{i=1}^n (\hat{A}_i - \bar{A})^2}{\sum_{i=1}^n (A_i - \bar{A})^2} \quad | \quad R^2 \in [0, 1] \quad (17)$$

where  $A_i$  is the value of the given raw mass isotopomer abundance from the  $i$ th measurement,  $\hat{A}_i$  the corresponding regression value from the best fitting MID, and  $\bar{A}$  is the mean of the raw mass isotopomer abundances across all measurements.

To put it less formally:  $R^2$  shows which fraction of the total variation in the measured values can be explained by the linear regression model. Ideally,

all variation is explained by the model ( $R^2 = 1$ ). In the worst case, none of the variation can be explained by the regression model ( $R^2 = 0$ ).

#### *2.4.3. Confidence intervals*

The previous measures only provide an overall quality measure for the whole MID. To judge the precision of the individual mass isotopomer abundances, another measure is required. When replicate measurements are provided, confidence intervals for the relative mass isotopomer abundances can be computed. The NTFD application reports 95% confidence intervals for the mass isotopomer abundances.

### **3. Practical considerations**

To obtain meaningful results from NTFD certain critical points need to be taken into account. Starting with important experimental considerations we will discuss how to avoid pitfalls in data analysis and highlight how common problems can be solved.

#### *3.1. Experimental considerations*

A prerequisite for a NTFD analysis is a stable isotope labeling experiment and a parallel experiment with the unlabeled tracer. To ensure that the maximum number of labeled compounds is detected, both experiments must be performed under the same conditions using the same cell line and growth medium.

Although NTFD can work on single measurements, the use of replicate measurements is strongly recommended. Replicates can reduce the effect of analytical variations and will make the MID calculation more robust. If

replicates are available, NTFD provides additional statistical quality measures for the calculated MIDs (subsection 2.4). Usually, three measurements per experiment provide sufficiently good results.

### *3.1.1. Choice of isotopic tracer*

For the design of a stable isotope labeling experiment, the proper choice and amount of tracer is crucial. This choice completely depends on the biological question asked [28]. For example, if the goal is to differentiate between metabolites produced and analytical background, all carbon substrates should be replaced by their fully labeled analogues. To completely label an organism a defined minimal medium, if it is available for the organism of interest, is most conveniently used.

Generally, isotopic enrichment from any isotopic element can be analyzed with NTFD, as long as there is no significant effect on  $M_2$  abundance, which would not be corrected for. However, special care has to be taken in case of deuterium. Deuteration is well-known to affect chromatographic retention, and different isotopologues may be separated completely. This can turn out problematic for two reasons: (1) the isotopologue mixture can be spread out over multiple deconvoluted mass spectra and (2) the matching of the labeled and unlabeled spectrum may not succeed due to shifts in retention time.

A combination of multiple tracers is generally possible, but will further complicate the already difficult interpretation of the resulting MIDs. In most cases it is advisable to distribute tracers among different experiments, except if the goal is to completely label an organism.

Once the tracer is chosen, the second important decision is the ratio at which it is introduced. On the one hand, a higher tracer abundance will lead

to higher and therefore better detection of isotopic enrichment in downstream metabolites. On the other hand, complete isotopic enrichment may lead to very dissimilar mass spectra from the labeled and unlabeled sample which cannot be matched properly. In such a case labeled compounds may be overseen, since the detection of isotopic enrichment by NTFD depends on the correct matching of these mass spectra. However, in most cases, NTFD works well with higher isotopic enrichment.

An important parameter for subsequent data analysis is the incubation time with the selected tracer. For many applications, it is desirable that the system reaches isotopic steady-state, meaning that isotopic labeling does not change over time. The required incubation time can range from several minutes to several days and needs to be determined from stable isotope labeling time-series of the specific system of interest. With *in vivo* labeling of complex organisms isotopic steady-state may not be achievable at all. However, isotopic non-steady-state does not affect the detection and quantification of isotopic enrichment *per se*, but needs to be accounted for during interpretation of the resulting MIDs.

### 3.1.2. GC-MS measurements

Ideally, the labeled and the unlabeled samples are measured subsequently on the same instrument. Standard GC-MS temperature programs and settings for the analysis of trimethylsilyl (TMS) derivatives of metabolites can be found here [29]. In case measurements are from different runs or instruments, retention time stability must be ensured. To aid compound matching, a calibration mixture such as an *n*-alkane mix (usually C<sub>10</sub> to C<sub>40</sub>) should be measured to calculate the Kováts retention index [30].

The usual metabolomics quality controls, like *e.g.* blank samples, should be included in the measurement as a quality control measure. If a compound that is isotopically enriched is also found to be present in any blank measurements, the respective MIDs need to be interpreted with caution, since the relative  $M_0$  abundance may be overestimated due to label dilution from the exogenous compound.

A similar problem is caused by detector saturation. If signal intensities lie outside the linear range of the detector, or if the detector is fully saturated, signal intensities are skewed or clipped. The relative mass isotopomer abundance determined by the respective ion may then be underestimated. This can, depending on the extent, strongly impact MID calculation for the respective fragments.

### 3.2. Data analysis

As with most non-targeted methods, NTFD data analysis can be tedious and frustrating if the wrong parameters are applied. In the following section, the most important steps and the respective parameters controlling a successful analysis are explained. The NTFD implementation is freely available for Linux and Windows operating systems. The software and detailed documentation of the user interface, along with sample data and a step-by-step tutorial are available at <http://ntfd.mit.edu>. Briefly, there are four different dialogs the user can interact with (Figure 4): (1) Selection of labeled measurements in netCDF or MetaboliteDetector [31] format. (2) Selection of unlabeled measurements in netCDF or MetaboliteDetector format. (3) The settings tab containing the important parameters for compound and label detection. (4) The result tab showing the list of detected labeled compounds

and their respective fragment MIDs.

### *3.2.1. Data preprocessing*

Before starting an NTFD analysis, it is important to check the quality of all chromatograms. The outcome of the analysis depends heavily on the correct preprocessing of the raw data. Before NTFD can start to detect isotopically enriched compounds, a peak picking and mass spectral deconvolution has to be performed. The raw data is scanned for peaks in the signal intensity which cross a certain threshold of absolute intensity and signal-to-noise ratio. Any signal below this threshold is discarded. All peaks detected at this stage are used for mass spectral deconvolution, that is they are grouped to specific mass spectra according to their retention time. A narrower deconvolution window will ensure that closely eluting peaks are separated, and a broader deconvolution window is necessary for broad peaks or for deuterated compounds where different isotopologues can elute at slightly different retention times.

This stage is very important, because it forms the basis on which all subsequent label detection will be performed. Therefore, the parameters should be well chosen. If the peak picking is too sensitive it will increase processing time and may yield an increased number of false positives. On the other hand this step needs to be sensitive enough to detect low-intensity isotopic peaks (Figure 5). To avoid a high number of false positives during the detection of isotopic enrichment, low quality spectra are best filtered out already after deconvolution. For example most EI-MS spectra with less than 20 peaks are usually of low quality. Although NTFD can directly process netCDF files, perform peak picking and spectrum deconvolution and the default parame-

ters will often fit, it is advisable to check the settings used within a dedicated analysis software such as MetaboliteDetector [31]. The advantage of MetaboliteDetector is that the results of these preprocessing steps can be visualized, therefore, effects of different parameters can be evaluated more easily. Moreover, MetaboliteDetector can calculate the Kováts retention index [30] and NTFD can directly use the generated files.

**Critical parameters controlling this step (Figure 4):**

**Peak threshold** The minimum signal-to-noise ratio for every chromatographic peak (default value: 5).

**Min Peak height** The peak threshold in absolute signal intensity (default value: 5).

**Deconvolution width** The number of scans by which chromatographic peaks are allowed to differ to still be considered as part of the same mass spectrum (default value: 5).

**Redetect all compounds** The peak detection and spectrum deconvolution step only has to be performed once. However, if any of the settings are updated, it has to be repeated for all measurements.

*3.2.2. Detection and quantification of isotopic enrichment*

While the detection of isotopically enriched fragments mainly depends on the correct peak picking and deconvolution parameters described above, the correct quantification additionally requires a correct  $\mathbf{M}_1$  correction (subsection 2.3).

**Critical parameters controlling this step (Figure 4):**

**M<sub>1</sub> correction** Needs to be adjusted for the isotopic element used in the tracer (Equation 7). Only A+1 elements are currently supported. This value is the ratio of the natural relative abundance of the heavy  $p_h$  and light  $p_l$  isotope times 100:

$$M1corr = \frac{p_h}{p_l}$$

Natural isotope abundances can be found in [32]. Most common correction values are:

$$\begin{aligned} M1corr_{^{13}\text{C}} &= \frac{p_{^{13}\text{C}}}{p_{^{12}\text{C}}} = \frac{0.0107}{0.9893} &&= 0.0108, \text{ for } ^{13}\text{C} \text{ labeling} \\ M1corr_{^{15}\text{N}} &= \frac{p_{^{15}\text{N}}}{p_{^{14}\text{N}}} = \frac{0.00364}{0.99636} &&= 0.00365, \text{ for } ^{15}\text{N} \text{ labeling} \\ M1corr_{^2\text{H}} &= \frac{p_{^2\text{H}}}{p_{^1\text{H}}} = \frac{0.000115}{0.999885} &&= 0.000115, \text{ for } ^2\text{H} \text{ labeling} \end{aligned}$$

### 3.2.3. Data filtering & quality control

One important part of data analysis is to exclude false positives from the result set. Although NTFD offers several different parameters to exclude false positively identified fragments or fragments with incorrectly calculated MIDs, it is absolutely essential to go back and consult the original data before interpreting the results. For example, fragment overlap, which impairs MID calculation, can be easily identified by inspecting the unlabeled mass spectrum. Furthermore, compounds which have not been detected as isotopically enriched need to be carefully checked before interpreting any negative results.

### Critical parameters controlling this step (Figure 4):

Tuning the following parameters will suffice for most applications. For

additional *Advanced* parameters Figure 4, which are described in the NTFD documentation, the default values should be used.

***Minimum  $R^2$***  Results can be filtered by their  $R^2$ . All fragments with a coefficient of determination below this threshold will be discarded.

***Maximal fragment deviation*** All fragments for which the sum of absolute values of relative mass isotopomer abundances deviates from 1 by more than this value will be discarded.

***Required number of labeled fragments*** Compounds with less labeled fragments will be discarded.

***Minimal and maximal isotopic enrichment*** Compounds for which the detected enrichment is very low are often false positives; moreover isotopic enrichment cannot be higher than the original tracer ratio. Results can be filtered for  $min \leq 1 - M_0 \leq max$ .

#### 3.2.4. Controlling sensitivity and specificity

The above described parameters allow the user to adjust the sensitivity and specificity of the label detection to the specific purpose of the study. For example, if the aim of the study is to validate active metabolic pathways, the information on isotopic enrichment *per se* is of more interest than the accuracy of the quantitative enrichment. In that case parameters should be optimized towards high sensitivity. Specifically the following settings could be applied: *Required number of labeled fragments* = 1, *Minimal isotopic enrichment* = 1,  $R^2 = 0.9$ , and *Maximal fragment deviation* =

0.2. As described above, it is of utter high importance with these sensitive settings to carefully check the results for false positives. However, if the goal of the study is to quantitatively compare MIDs between two conditions, settings should be stricter. As a rule of thumb, the  $R^2$  should be above 0.98 and the sum of absolute values should be less than 1.02 for MID interpretation. To get reliable MIDs the following settings could be used: *Required number of labeled fragments* = 2, *Minimal isotopic enrichment* = 5,  $R^2$  = 0.98, and *Maximal fragment deviation* = 0.02. Nevertheless, it is advisable to start with sensitive settings and optimize towards the best trade-off between specificity and sensitivity.

#### 4. NTFD applications

Most tools for the non-targeted detection of stable isotope labeling have been developed only recently. Therefore, it is not surprising that there have not been many non-targeted stable isotope labeling analyses yet. We will present previously published as well as additional potential applications of non-targeted detection of isotope labeling.

##### 4.1. Detecting genuine metabolites

In non-targeted metabolomics one is often confronted with a large number of chromatographic - mass spectrometric features where it is unclear whether these are artifacts from sample preparation, compounds from complex growth media, or native metabolites. After cultivating an organism on fully  $^{13}\text{C}$  enriched substrates, one can be sure that any carbon-containing compound that is not isotopically enriched is not produced by the organism but is of exogenous origin. Therefore, non-targeted detection of stable isotope labeling

has most commonly been applied to distinguish metabolites from analytical background and to determine mass spectrometric ions originating from the labeled and the unlabeled compound to use them for relative quantification of metabolite levels [33–37].

#### 4.2. $^{15}\text{N}$ tracing

Not only  $^{13}\text{C}$ , but also  $^{15}\text{N}$  labeling can provide interesting insights into metabolism. For mammalian cells, glutamine and glutamic acid are the main nitrogen sources. Therefore,  $[\text{U-}^{15}\text{N}]\text{glutamine}$  can be applied as a tracer to label most nitrogen containing metabolites in mammalian cells, which can in turn be detected with NTFD. In a more quantitative approach, Gaglio *et al.* [38] analyzed MIDs after  $[\alpha\text{-}^{15}\text{N}]\text{glutamine}$  labeling in a non-targeted manner to reveal altered glutamine utilization in K-Ras transformed cells.

#### 4.3. Analysis of mass spectrometric fragmentation

The non-targeted detection of stable isotope labeling in mass spectrometric fragments can be interesting for compounds with unknown fragmentation pathways. MIDs of individual fragments provide valuable information on elemental composition and origin of the respective atoms. Moreover, the MID can reveal overlapping fragments of the same mass [9]. For this purpose, we embedded the NTFD algorithm into the Fragment Formula Calculator (FFC) software [9]. FFC determines elemental composition and potential substructures of mass spectrometric fragment ions by a graph-based combinatorial approach. It can include information from stable isotope labeled spectra to rule out certain combinations.

#### *4.4. Metabolism of xenobiotics*

In pharmaceutical research, absorption, distribution, metabolism, and excretion (ADME) studies are conducted to determine the fate of a drug or other xenobiotics within an organism. Such studies often use radiolabeled compounds or analyze stable isotope labeling in a targeted manner [39]. NTFD can provide insights into metabolism of xenobiotics without the need of radiolabeled compounds. After applying a stable isotope labeled drug to an organism, NTFD can be used to detect all measured drug metabolites in a non-targeted manner.

#### *4.5. Model validation*

Genome-scale metabolic networks [40] are used for the stoichiometric modeling of metabolism to infer intracellular metabolic fluxes [41]. These metabolic networks are generated from transcriptomics or proteomics data. However, the presence of a certain mRNA or even protein does not tell anything about actual enzymatic activities. To this end, non-targeted tracer fate analyses can be used to determine metabolic pathways that are active under any given conditions. Furthermore, NTFD can be used to test the comprehensiveness of a metabolic network reconstruction. If a reconstruction is comprehensive, all isotopically enriched compounds should be identifiable. Otherwise, additional compounds need to be included.

#### *4.6. Tracing functional groups*

Dedicated tracers can be used in combination with NTFD to trace functional groups through metabolism or detect products of specific types of reactions. Methylation reactions and associated one-carbon metabolism are

important for many anabolic processes as well as for detoxification of many xenobiotics and inactivation of endogenous substances such as catecholamines [42]. The methyl groups are derived directly or indirectly from methyltetrahydrofolate. The main contributor to this pool is serine, or more specifically C-3 of serine. Therefore, [3- $^{13}\text{C}$ ]serine can be applied to label the one-carbon pool and all methylation products.

#### 4.7. Cofactor tracing

Lewis *et al.* [41] and Fan *et al.* [43] used deuterated glucose tracers to selectively label the cellular NADH or NADPH pool. [3- $^2\text{H}$ ]glucose specifically labels NADPH that is generated in the pentose phosphate pathway. Other glucose molecules that are undergoing glycolysis and are not entering the pentose phosphate pathway will lose the deuterium label to water. The deuterium from the labeled NADPH will be transferred as hydride ion by NADPH-dependent enzymes, mostly reductases and dehydrogenases, to their substrates. These products can then be detected by NTFD in a non-targeted manner. Analogously, [4- $^2\text{H}$ ]glucose can be used to label the cellular NADH pool and to trace NADH-dependent reductions [41].

#### 4.8. Non-targeted metabolic turnover analysis

Non-targeted metabolic turnover analysis, was presented by Nakayama *et al.* [44] to detect metabolic pathways affected by external perturbations. To this end, they performed principal component analysis (PCA) on a time-series of relative mass isotopomer abundances. Such analyses can reveal metabolic vicinity and group unidentified compounds into discrete pathways.

More such novel workflows are needed to make full use of the information contained in metabolome-wide stable isotope labeling datasets.

## 5. Summary & Outlook

The field of non-targeted stable isotope labeling analysis is still very young and algorithms and tools for the automated metabolome-wide detection of isotopic enrichment have only become available in recent years. Here we presented the NTFD algorithm, explained how NTFD can be used to determine isotopic enrichment in a non-targeted manner, and finally highlighted potential applications. The advancement of current tools and the development of novel workflows will help to benefit from the high potential of non-targeted stable isotope labeling data to provide further insights into the structure and dynamics of metabolic networks.

## 6. Acknowledgement

This project was supported by the Fonds National de la Recherche (FNR) Luxembourg (ATTRACT A10/03).

- [1] Chokkathukalam, A.; Kim, D.-H.; Barrett, M. P.; Breitling, R.; Creek, D. J. *Bioanalysis* **2014**, *6*, 511–524, DOI: 10.4155/bio.13.348.
- [2] Missios, P. *et al.* *Nat Commun* **2014**, *5*, 4924, DOI: 10.1038/ncomms5924.
- [3] Bodamer, O. A.; Halliday, D. *Arch Dis Child* **2001**, *84*, 444–448,

- [4] Metallo, C. M.; Gameiro, P. A.; Bell, E. L.; Mattaini, K. R.; Yang, J.; Hiller, K.; Jewell, C. M.; Johnson, Z. R.; Irvine, D. J.; Guarente, L.; Kelleher, J. K.; Heiden, M. G. V.; Iliopoulos, O.; Stephanopoulos, G. *Nature* **2012**, *481*, 380–384, DOI: 10.1038/nature10602.
- [5] Vallino, J. J.; Stephanopoulos, G. *Biotechnol Bioeng* **2000**, *67*, 872–885,
- [6] Yoo, H.; Stephanopoulos, G.; Kelleher, J. K. *J Lipid Res* **2004**, *45*, 1324–1332, DOI: 10.1194/jlr.M400031-JLR200.
- [7] Sellers, K.; Fox, M. P.; Bousamra, M., 2nd; Slone, S. P.; Higashi, R. M.; Miller, D. M.; Wang, Y.; Yan, J.; Yuneva, M. O.; Deshpande, R.; Lane, A. N.; Fan, T. W.-M. *J Clin Invest* **2015**, DOI: 10.1172/JCI72873.
- [8] Pickup, J. F.; McPherson, K. *Analytical Chemistry* **1976**, *48*, 1885–1890, DOI: 10.1021/ac50007a019.
- [9] Wegner, A.; Weindl, D.; Jäger, C.; Sapcariu, S. C.; Dong, X.; Stephanopoulos, G.; Hiller, K. *Anal Chem* **2014**, *86*, 2221–2228, DOI: 10.1021/ac403879d.
- [10] Dang, L. *et al. Nature* **2009**, *462*, 739–744, DOI: 10.1038/nature08617.
- [11] Michelucci, A.; Cordes, T.; Ghelfi, J.; Pailot, A.; Reiling, N.; Goldmann, O.; Binz, T.; Wegner, A.; Tallam, A.; Rausell, A.; Buttini, M.; Linster, C. L.; Medina, E.; Balling, R.; Hiller, K. *Proc Natl Acad Sci U S A* **2013**, *110*, 7820–7825, DOI: 10.1073/pnas.1218599110.
- [12] Hiller, K.; Metallo, C. M.; Kelleher, J. K.; Stephanopoulos, G. *Anal Chem* **2010**, *82*, 6621–6628, DOI: 10.1021/ac1011574.

- [13] Hiller, K.; Wegner, A.; Weindl, D.; Cordes, T.; Metallo, C. M.; Kelleher, J. K.; Stephanopoulos, G. *Bioinformatics* **2013**, *29*, 1226–1228, DOI: 10.1093/bioinformatics/btt119.
- [14] Huang, X.; Chen, Y.-J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J. *Anal Chem* **2014**, DOI: 10.1021/ac403384n.
- [15] Creek, D. J.; Chokkathukalam, A.; Jankevics, A.; Burgess, K. E. V.; Breitling, R.; Barrett, M. P. *Anal Chem* **2012**, *84*, 8442–8447, DOI: 10.1021/ac3018795.
- [16] Bueschl, C.; Kluger, B.; Berthiller, F.; Lirk, G.; Winkler, S.; Krska, R.; Schuhmacher, R. *Bioinformatics* **2012**, DOI: 10.1093/bioinformatics/bts012.
- [17] Cho, K.; Mahieu, N.; Ivanisevic, J.; Uritboonthai, W.; Chen, Y.-J.; Siuzdak, G.; Patti, G. J. *Anal Chem* **2014**, *86*, 9358–9361, DOI: 10.1021/ac5029177.
- [18] Chokkathukalam, A.; Jankevics, A.; Creek, D. J.; Achcar, F.; Barrett, M. P.; Breitling, R. *Bioinformatics* **2013**, *29*, 281–283, DOI: 10.1093/bioinformatics/bts674.
- [19] Chace, D. H.; Abramson, F. P. *Anal Chem* **1989**, *61*, 2724–2730,
- [20] Sauer, U. *Mol Syst Biol* **2006**, *2*, 62, DOI: 10.1038/msb4100109.
- [21] Niedenführ, S.; Wiechert, W.; Nöh, K. *Current Opinion in Biotechnology* **2015**, *34*, 82 – 90, DOI: <http://dx.doi.org/10.1016/j.copbio.2014.12.003>, Systems Biology – Nanobiotechnology.

- [22] IUPAC, *Gold Book—Compendium of Chemical Terminology*; International Union of Pure and Applied Chemistry, 2012; Version 2.3.2 2012-08-19.
- [23] Hellerstein, M. K.; Neese, R. A. *Am J Physiol* **1999**, *276*, E1146–E1170,
- [24] Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. *Pure and Applied Chemistry* **2013**, *86*, 1515–1609, DOI: 10.1351/PAC-REC-06-04-06.
- [25] Christensen, B.; Nielsen, J. *Metab Eng* **1999**, *1*, 282–290, DOI: 10.1006/mben.1999.0117.
- [26] Stein, S. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 770 – 781, DOI: 10.1016/S1044-0305(99)00047-1.
- [27] Jennings, M. E.; Matthews, D. E. *Anal Chem* **2005**, *77*, 6435–6444, DOI: 10.1021/ac0509354.
- [28] Metallo, C. M.; Walther, J. L.; Stephanopoulos, G. *J Biotechnol* **2009**, *144*, 167–174, DOI: 10.1016/j.jbiotec.2009.07.010.
- [29] Sapcariu, S. C.; Kanashova, T.; Weindl, D.; Ghelfi, J.; Dittmar, G.; Hiller, K. *MethodsX* **2014**, *1*, 74 – 80, DOI: 10.1016/j.mex.2014.07.002.
- [30] Kováts, E. *Helvetica Chimica Acta* **1958**, *41*, 1915–1932, DOI: 10.1002/hlca.19580410703.
- [31] Hiller, K.; Hangebrauk, J.; Jäger, C.; Spura, J.; Schreiber, K.; Schomburg, D. *Anal Chem* **2009**, *81*, 3429–3439, DOI: 10.1021/ac802689c.

- [32] Berglund, M.; Wieser, M. E. *Pure Appl. Chem* **2011**, *83*, 397–410,
- [33] Weindl, D.; Wegner, A.; Jäger, C.; Hiller, K. *J Chromatogr A* **2015**, *1389*, 112–119, DOI: 10.1016/j.chroma.2015.02.025.
- [34] Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. *Anal Chem* **2014**, *86*, 4675–4679, DOI: 10.1021/ac5009089.
- [35] de Jong, F. A.; Beecher, C. *Bioanalysis* **2012**, *4*, 2303–2314, DOI: 10.4155/bio.12.202.
- [36] Bueschl, C.; Kluger, B.; Lemmens, M.; Adam, G.; Wiesenberg, G.; Maschietto, V.; Marocco, A.; Strauss, J.; Bödi, S.; Thallinger, G. G.; Krska, R.; Schuhmacher, R. *Metabolomics* **2013**, 1–16, DOI: 10.1007/s11306-013-0611-0.
- [37] Yang, S.; Hoggard, J. C.; Lidstrom, M. E.; Synovec, R. E. *J Chromatogr A* **2013**, *1317*, 175–185, DOI: 10.1016/j.chroma.2013.08.059.
- [38] Gaglio, D.; Metallo, C. M.; Gameiro, P. A.; Hiller, K.; Danna, L. S.; Balestrieri, C.; Alberghina, L.; Stephanopoulos, G.; Chiaradonna, F. *Mol Syst Biol* **2011**, *7*, 523, DOI: 10.1038/msb.2011.56.
- [39] Mutlib, A. E. *Chem Res Toxicol* **2008**, *21*, 1672–1689, DOI: 10.1021/tx800139z.
- [40] Thiele, I. *et al. Nat Biotechnol* **2013**, *31*, 419–425, DOI: 10.1038/nbt.2488.

- [41] Lewis, C. A.; Parker, S. J.; Fiske, B. P.; McCloskey, D.; Gui, D. Y.; Green, C. R.; Vokes, N. I.; Feist, A. M.; Vander Heiden, M. G.; Metallo, C. M. *Mol Cell* **2014**, *55*, 253–263, DOI: 10.1016/j.molcel.2014.05.008.
- [42] Meiser, J.; Weindl, D.; Hiller, K. *Cell Commun Signal* **2013**, *11*, 34, DOI: 10.1186/1478-811X-11-34.
- [43] Fan, J.; Ye, J.; Kamphorst, J. J.; Shlomi, T.; Thompson, C. B.; Rabinowitz, J. D. *Nature* **2014**, *510*, 298–302, DOI: 10.1038/nature13236.
- [44] Nakayama, Y.; Tamada, Y.; Tsugawa, H.; Bamba, T.; Fukusaki, E. *Metabolites* **2014**, *4*, 722–739, DOI: 10.3390/metabo4030722.

## 7. Figure Legends

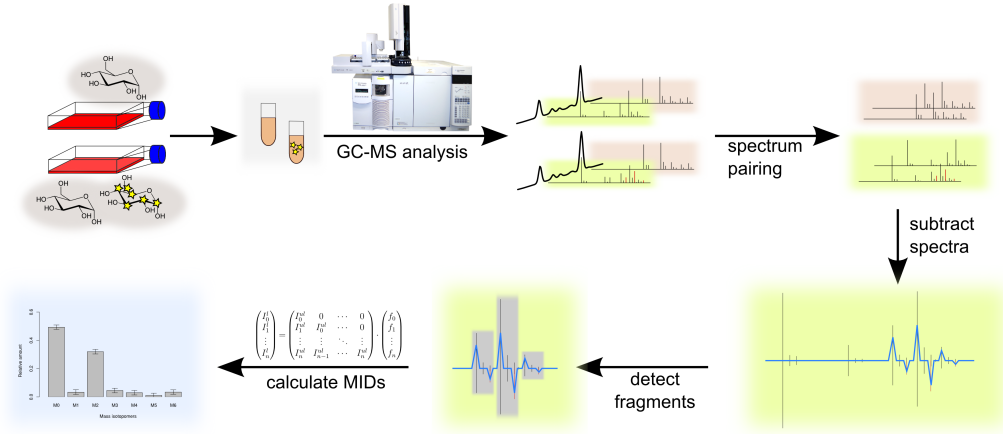


Figure 1: The NTFD algorithm operates on GC-EI-MS measurements of an isotopically enriched and non-enriched sample. For each compound, mass spectra from both measurements are paired. Isotopically enriched fragments are determined from characteristic patterns in the difference spectrum. For each of these enriched fragments the MID is determined.

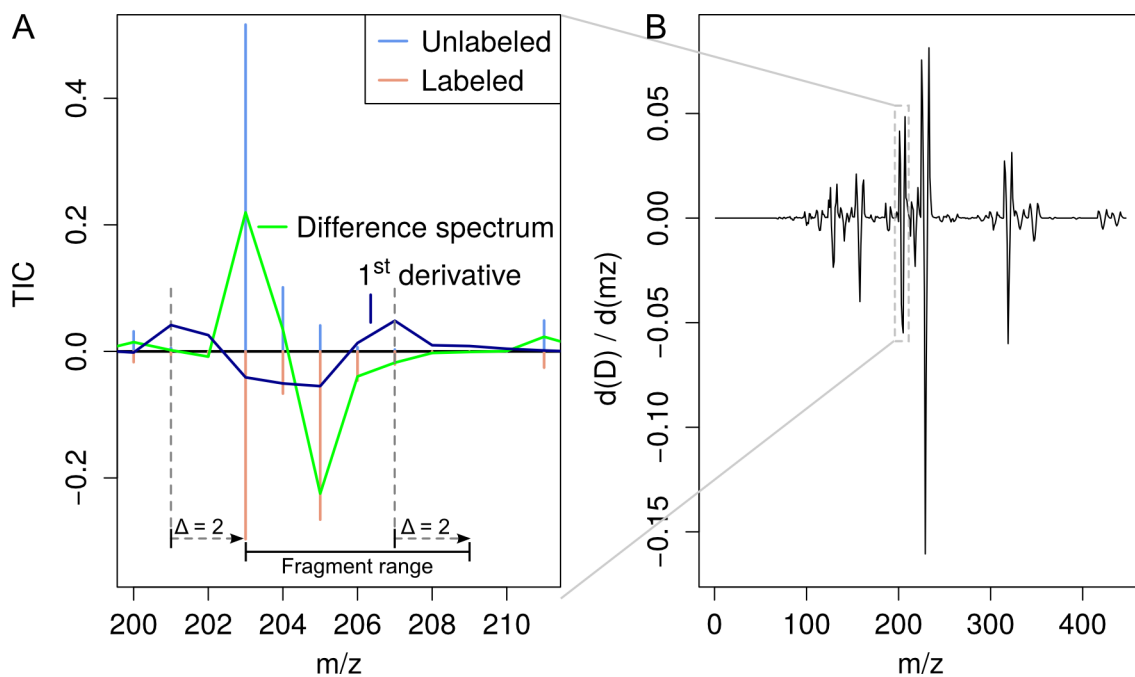


Figure 2: NTFD detects labeled fragments in the difference spectrum of mass spectra derived from a potentially isotopically enriched and non-enriched compound.

A: Mass spectrum of a fragment of the unlabeled compound and its labeled analogue, normalized to total ion current (TIC). Isotopic enrichment leads to characteristic peak patterns in the difference spectrum and its first derivative. Fragment boundaries are detected in the first derivative of the smoothed difference spectrum. The boundaries are shifted by 2 units due to the applied smoothing algorithm.

B: First derivative ( $\frac{dD}{d(m/z)}$ ) of the full difference spectrum of glutamine 4TMS. Two positive peaks, separated by a negative peak, mark an isotopically enriched fragment.

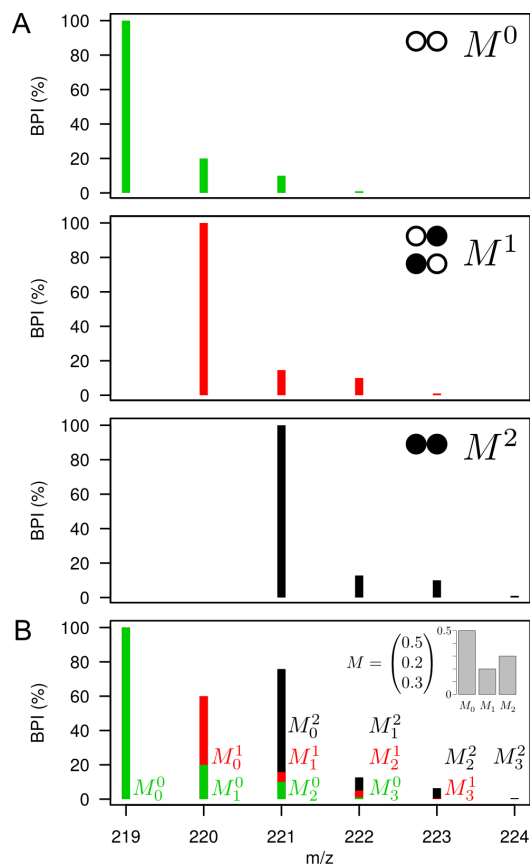


Figure 3: Artificial isotopic enrichment and natural isotope abundance determine isotopic peak patterns in a mass spectrum after stable isotope labeling experiments. Stable isotope labeling of a compound with two enrichable positions.

A: The mass spectra of all three mass isotopomers that can arise from stable isotope labeling show isotopic peaks due to the natural isotope abundance in the non-enriched positions. The natural isotope contribution, *i.e.* the relative intensity of the M+1 peak, decreases with increasing artificial enrichment. The relative M+0 intensity increases by the same value. Filled circle: heavy isotope; empty circle: light isotope. Signal intensity is scaled to base peak intensity (BPI).

B: After a stable isotope labeling experiment the mass spectrum of the given compound is a mixture of the natural MID of the three artificial mass isotopomers. This spectrum is the average of  $M^0$ ,  $M^1$  and  $M^2$ , weighted by the corrected MID  $M$ .

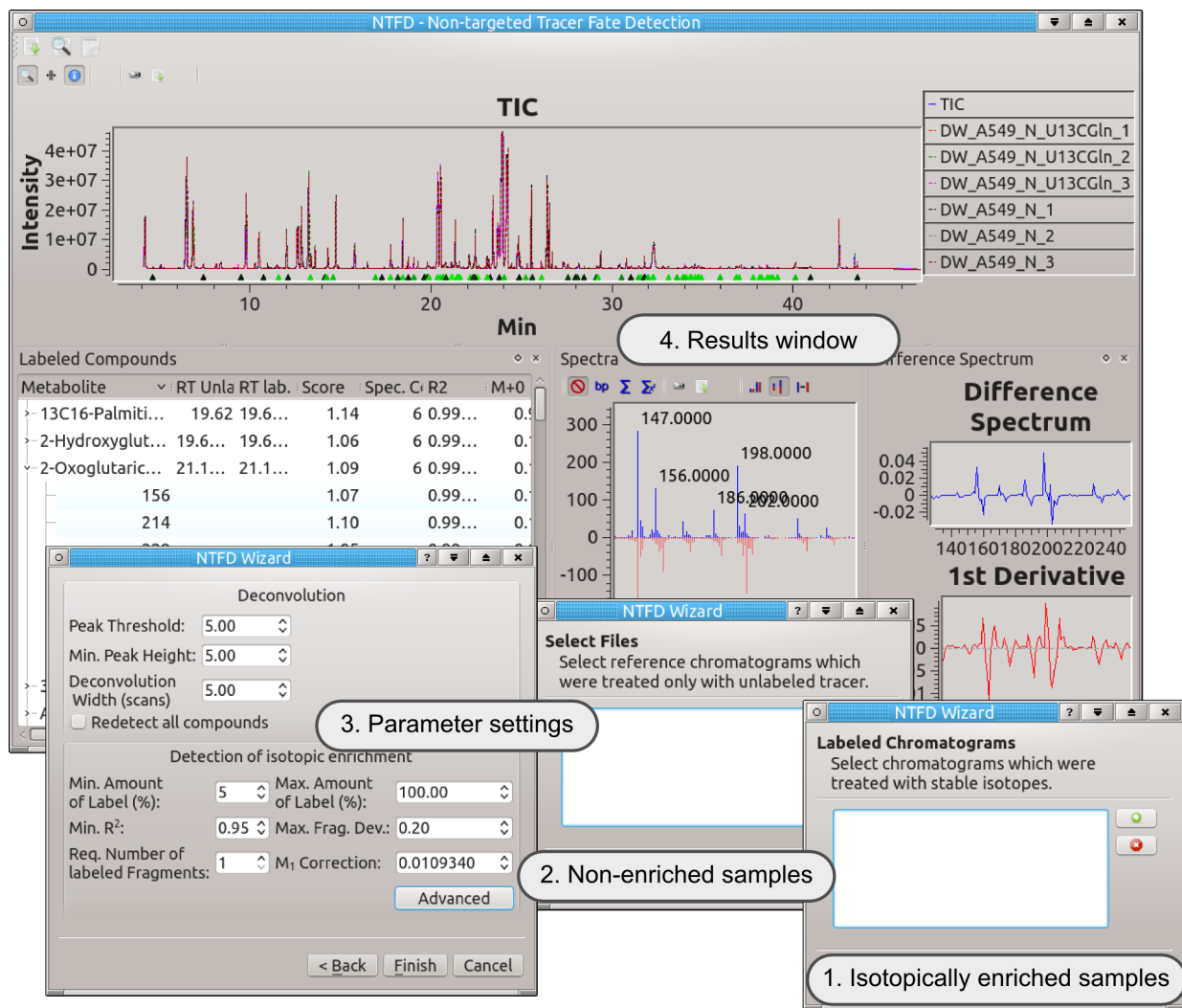


Figure 4: Graphical user interface of the NTFD implementation [12]. The program is freely available for Linux and Windows operating systems. After GC-MS measurement, (1) the files of the isotopically enriched and (2) the non-enriched samples need to be selected, (3) Parameters for the detection and quantification of isotopic enrichment need to be adjusted, then (4) NTFD will determine all isotopically enriched compounds and fragments along with their MIDs.

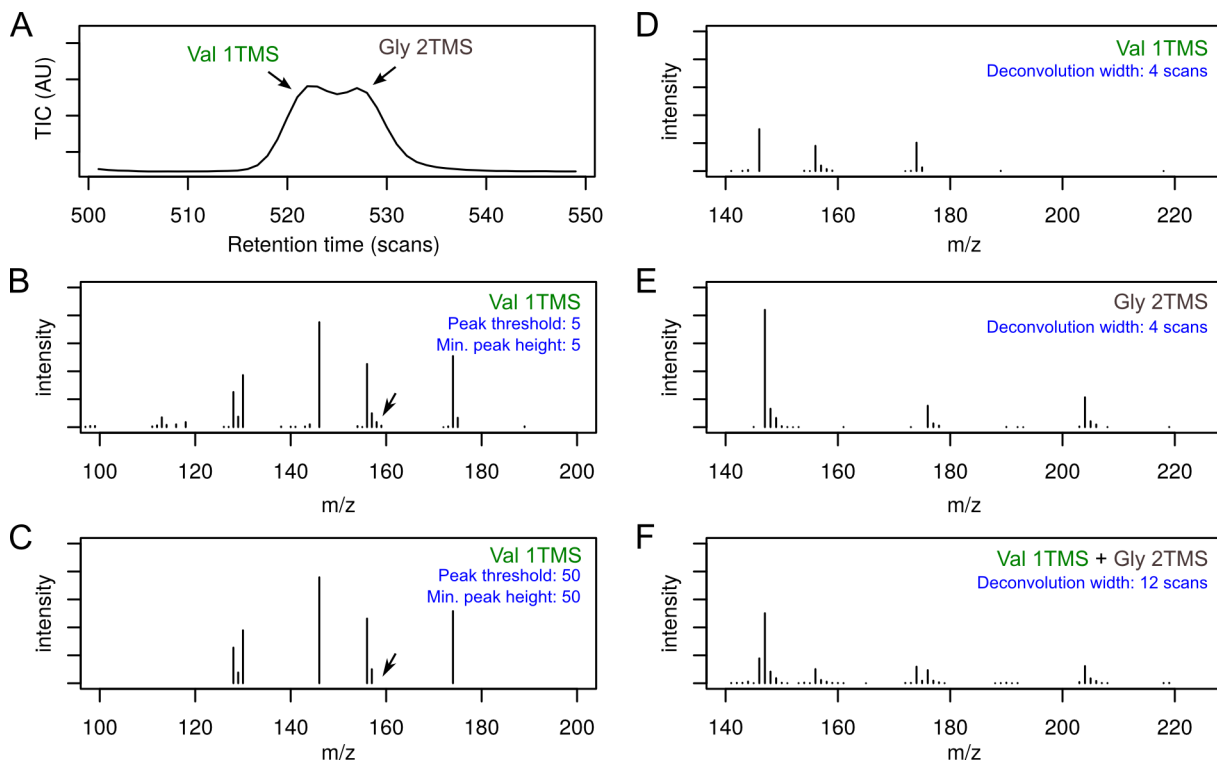


Figure 5: Proper peak picking and mass spectrum deconvolution is a prerequisite for meaningful NTFD results.

B,C: Insensitive peak detection will miss isotopic peaks or whole fragments and impair calculated MIDs.

D-F: Deconvolution settings need to be adjusted to properly extract pure mass spectra (D,E) of close-by chromatographic peaks of coeluting compounds (A). If deconvolution width is set too high, close-by chromatographic peaks will be assigned incorrectly to a single compound (F) and overlapping fragments will give rise to misleading MIDs.

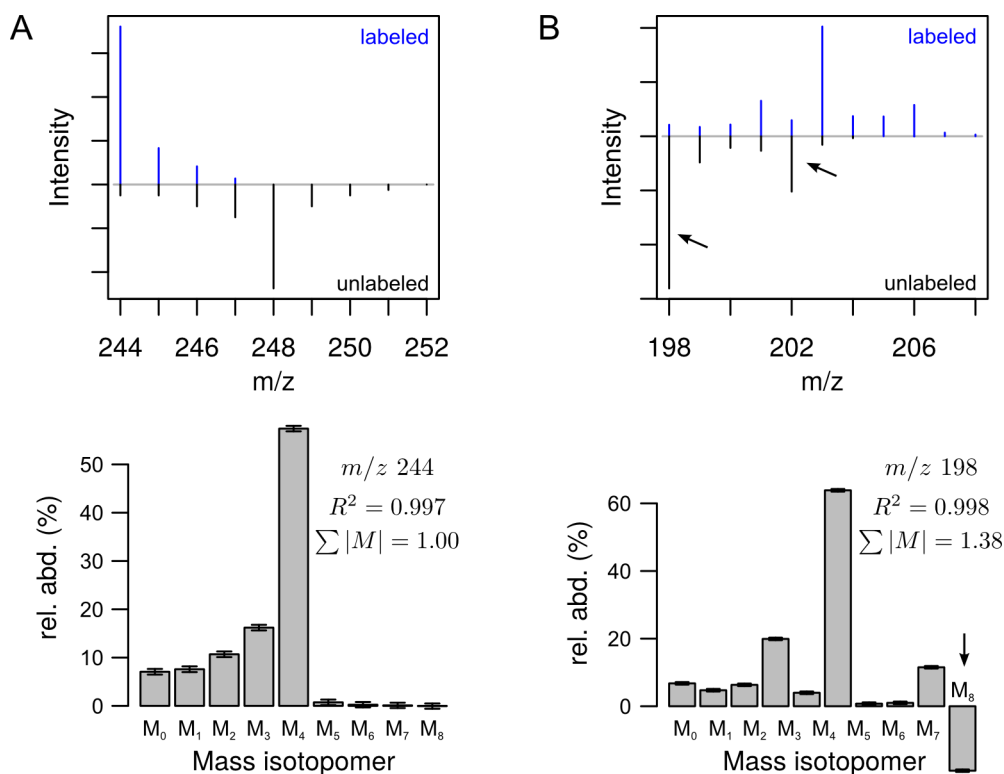


Figure 6: Quality measures for MID determination. MIDs of 2-oxoglutaric acid 2TMS 1MeOX determined after incubating mammalian cells with [U- $^{13}\text{C}$ ]glutamine.

A: Fragment  $m/z$  244: High quality mass spectra without fragment overlap yield well determined MIDs (small confidence intervals,  $R^2 \approx 1.00$ ,  $\sum |M| \approx 1.00$ ).

B: Overlapping fragments impair calculated MIDs. Fragments  $m/z$  198  $m/z$  202 are overlapping causing a high negative  $M_9$  abundance in the determined MID. Such MIDs are unusable, despite low confidence intervals and high  $R^2$ .