NTFD Documentation

Thekla Cordes, Andre Wegner, Daniel Weindl, Karsten Hiller*

December 2012

This document provides a basic overview over NTFD functionality. It is organized in two parts: a user manual and a step-by-step tutorial. The first part describes the graphical user interface (GUI), how to import data for analysis and important settings. The second part is a step-by-step tutorial that shows the complete analysis of a sample dataset.

Contents

1 Introduction **2**

2 User manual **2**

2.1 Installation 2
2.2 Graphical user interface (GUI) 2
2.3 Importing data 7
2.4 Settings 7
2.5 Recommended Settings 8

3 Step-by-step tutorial **9**

3.1 Start program and load samples 9
3.2 Set the right settings 11
3.3 Inspecting the results 11

4 References **14**

*mailto:karsten.hiller@uni.lu

1 Introduction

NTFD is a Qt4 based software package that implements a method for the non-targeted detection of stable isotope labeled tracer derived compounds [2]. NTFD will detect all compounds derived from a stable isotope labeled tracer in a set of GC/MS chromatograms and present their corresponding mass isotopomer distributions (MIDs). The software is implemented in the object orientated language C++ and comes with an easy-to-use graphical user interface. It is freely available for Debian and Redhat based Linux distributions as well as for Windows.

2 User manual

2.1 Installation

NTFD packages can be downloaded from `http://ntfd.mit.edu/index.php/download` either as deb or rpm for Linux or with an installer for windows.

2.1.1 Linux

Open a terminal and change to the directory containing the downloaded file. Then type the following in the command line:

`rpm -i filename` for rpm based systems

`sudo dpkg-i filename.deb` for Debian based systems

where filename is the name of the deb or rpm file (e.g `ntfd-1.0-Linux.deb`).

2.1.2 Windows

Start the windows installer and follow the instructions.

2.2 Graphical user interface (GUI)

The graphical user interface is organized in one mainwindow containing four subwindows.

2.2.1 Main window

The mainwindow toolbar (Fig. 1) provides the following functionality:

Export results to tsv Click this button to export the data shown in the labeled compounds window (see 2.2.3). The exported data can be opened by common spreadsheet programs like OpenOffice Calc or Microsoft Excel.

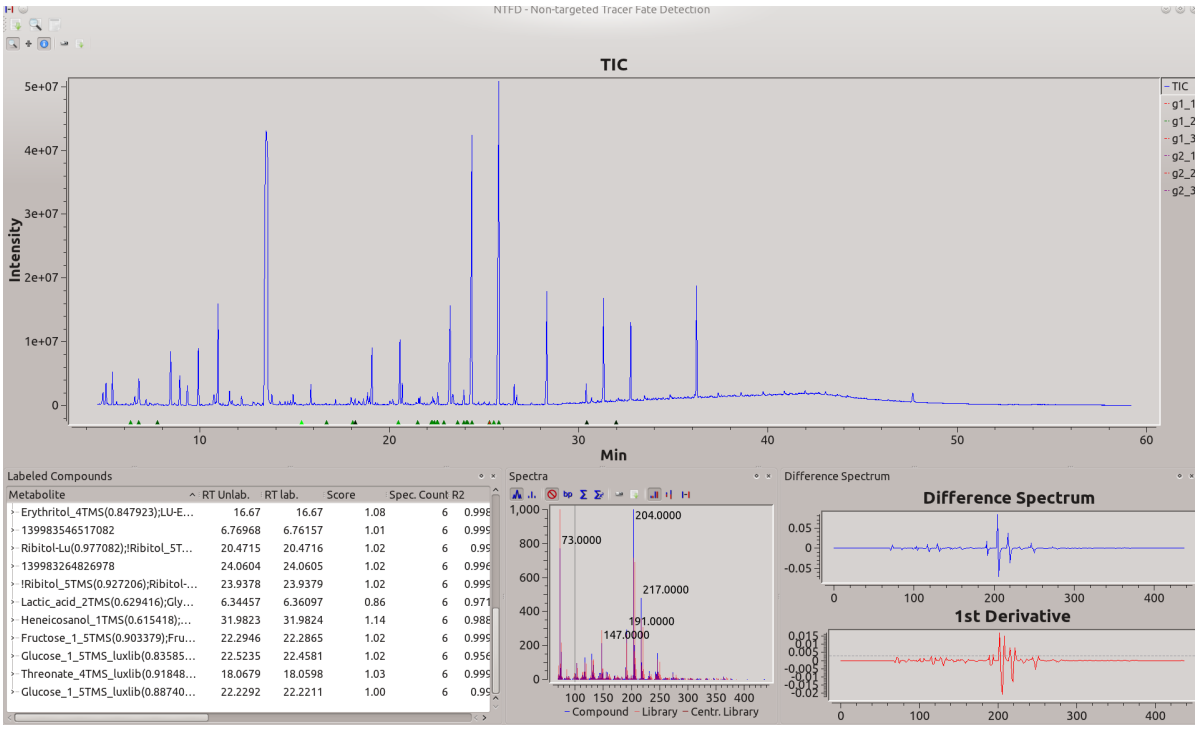

Figure 1: NTFD Main window

Identify compounds If this button is clicked, NTFD tries to identify the compounds shown in the labeled compounds window (see 2.2.3) based on the provided reference library. Identification is based on a spectrum similarity score (normalized dot product). In case retention indices are available, the spectrum similarity score is combined with a retention index similarity score.

Generate MetaboliteDetector library from detected compounds If this button is clicked, a MetaboliteDetector [1] library either for the unlabeled or labeled spectra is generated.

2.2.2 TIC Chromatogram

The Total Ion Current chromatogram (TIC) (Fig. 2) is displayed in the upper part of the user interface of NTFD.

The toolbar provides the following functionality:

Zoom If the zoom button is checked the following operations can be performed:

- Left button: Holding the left button and dragging the mouse zooms into the selected part of the chromatogram.

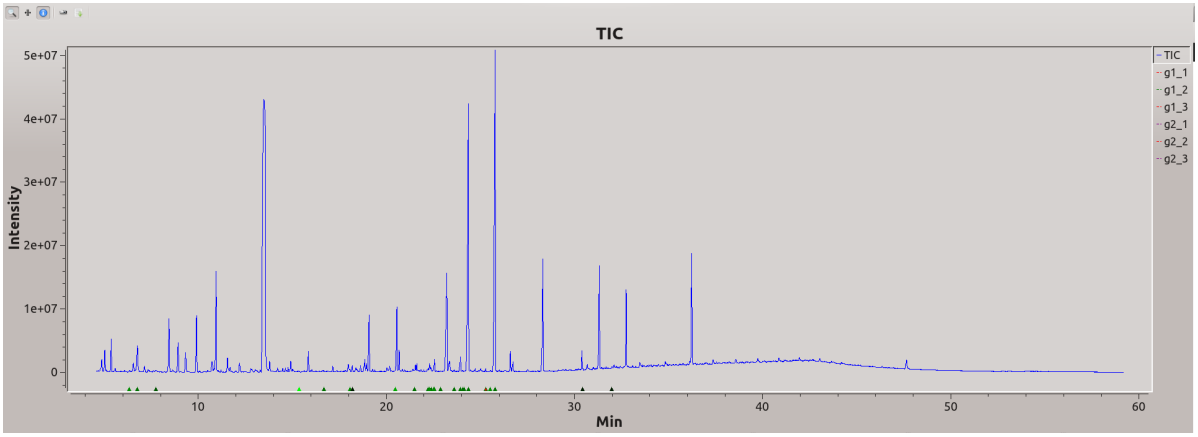

Figure 2: TIC chromatogram view

- Right button: Clicking the right button zooms out one step.
- Middle button: Holding the middle button and dragging the mouse moves the displayed chromatogram.

Component info Labeled compounds are marked by colored triangeles.

Save image Click this button to export the displayed TIC view as an image. NTFD supports many image formats including PNG, GIF, TIF and JPG. The export format as well as the desired image size can be chosen in the "Image Export" dialog.

tsv report Click this button to export the displayed TIC view as tsv (tab-separated values) data. This data can be opened by most spreadsheet programs like OpenOffice Calc or Excel.

2.2.3 Labeled Compounds

Labeled Compounds

Metabolite	RT Unlab.	RT lab.	Score	Spec. Count	R2	M+0	M+1	M+2	M+3	M+4	M+5	M+6
Ribitol_xTMS-luxlib(0.985931);Ma...	20.4715	20.4716	1.02	6	0.996998	0.38	0.07	0.14	0.11	0.29	0.01	
Glucose_1_5TMS_luxlib(0.767637);...	22.8668	22.8669	1.04	6	0.990168	0.52	0.02	0.04	0.44	*-0.00*	*-0.00*	
Mannose_1_1MeOX_5TMS_luxlib(...	24.0604	24.0605	1.02	6	0.996867	0.53	*0.00*	0.01	0.02	0.44	*-0.00*	
Glutamic_acid_3TMS_luxlib(0.7860...	21.477	21.4771	1.06	6	0.98945	0.51	*0.01*	*-0.02*	*0.01*	0.45	*-0.01*	
Glucose_1_5TMS_luxlib(0.827132);...	22.4581	22.4581	1.06	6	0.982444	0.53	*-0.02*	0.48	*0.01*	*-0.00*	*0.00*	
LU-Erythrose-Std1(0.81593);LU-Ery...	16.67	16.67	1.08	6	0.998395	0.54	0.02	0.02	0.41	0.04	*0.00*	
Glucose_2_5TMS_luxlib(0.847882);...	26.047	26.0471	1.01	6	0.991854	0.52	0.06	0.40	*-0.00*	0.01	*-0.00*	
Mannose_1_1MeOX_5TMS_luxlib(...	23.9378	23.9379	1.02	6	0.999864	0.51	0.00	0.00	0.02	0.47	*-0.00*	
Glucose_1_5TMS_luxlib(0.961538);...	24.3548	24.3548	1.06	6	0.997893	0.53	0.01	0.02	0.45	0.03	*-0.00*	
Glucose_2_5TMS_luxlib(0.966912);...	25.7854	25.7855	1.06	6	0.997944	0.52	0.01	0.02	0.46	0.03	*-0.00*	

Figure 3: List of labeled compounds.

The results (Fig. 3) are displayed in the lower left part of NTFD's user interface. For all labeled compounds the mass isotopomer distribution, coefficient of determination (R^2), spectrum count (number of files in which the compound was detected), retention time of the unlabeled compound, retention time of the labeled compound and the name of the compound (after identification, otherwise a number is displayed). These data can be exported in a tab-separated value (tsv) format by clicking the tsv export button from the main window's toolbar (see 2.2.1).

2.2.4 Spectra

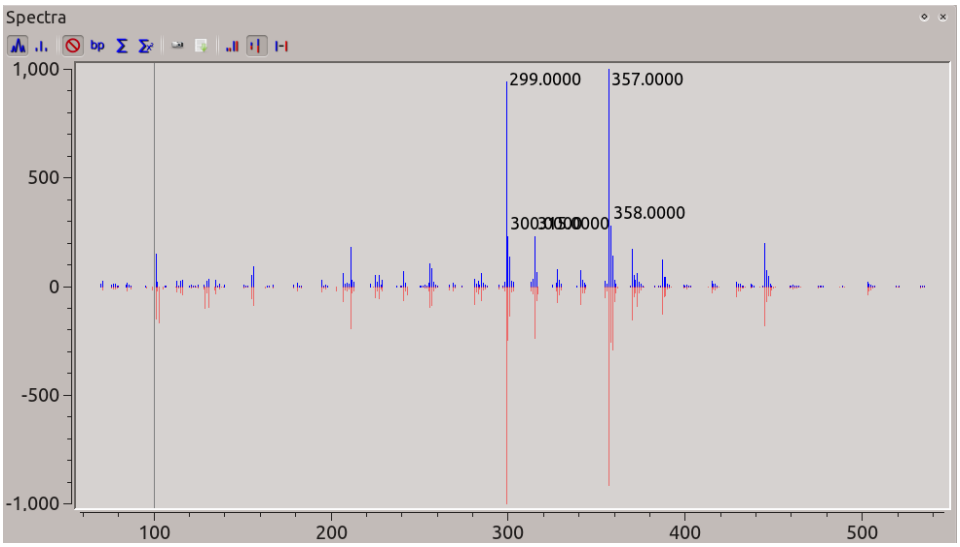

Figure 4: Mass spectrum view.

The mass spectrum of the labeled (in red) and unlabeled (in blue) compound selected in the chromatogram view is shown in the lower middle part of the NTFD user interface (Fig. 4).

The following display options can be selected:

Show profile spectrum Show profile spectrum.

Show centroided spectrum Show centroided spectrum.

No normalization Show original data (non-normalized).

Base peak normalization Show intensities relative to the base peak.

Relative intensity normalization Normalize to total ion current (TIC).

Unit normalization Scale to unit variance.

Normal mode Both, labeled and unlabeled spectrum are displayed in positive direction.

Split mode Unlabeled spectrum is displayed in positive direction and the labeled spectrum in negative direction.

Difference mode The difference spectrum is shown, unlabeled spectrum in positive direction and the labeled spectrum in negative direction.

Save spectrum as image Click this button to export the displayed spectra view as an image. NTFD supports many image formats including PNG, GIF, TIF and JPEG. The export format as well as the desired image size can be chosen in the "Image Export" dialog.

tsv report Click this button to export the masses and intensities of the currently displayed spectra as tsv data. This data can be opened by common spreadsheet programs like OpenOffice Calc or Microsoft Excel.

2.2.5 Difference spectrum

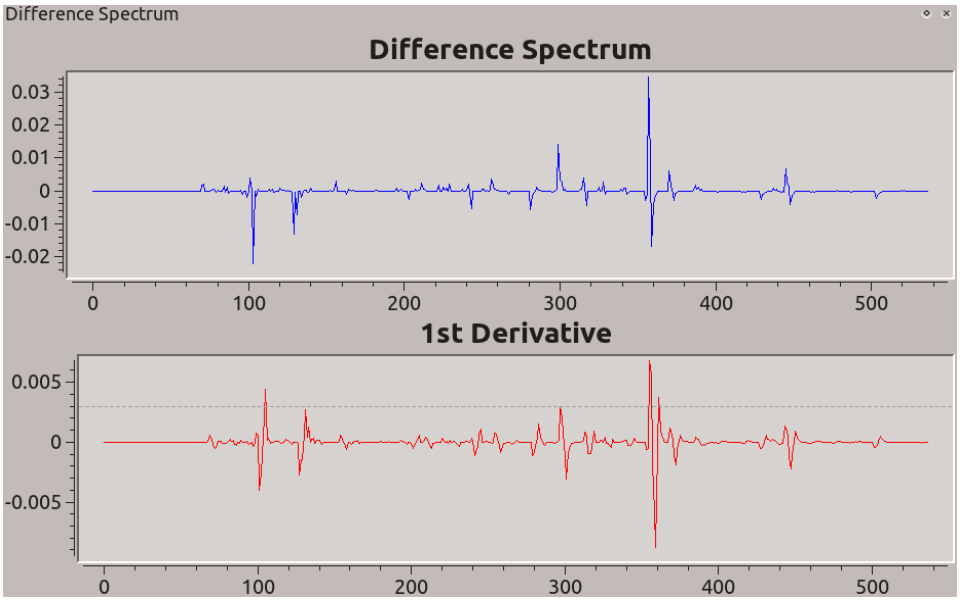

Figure 5: Difference spectrum view. Difference spectrum and first derivative.

The difference spectrum and the first derivative are displayed in the lower right part of the user interface (Fig. 5). The difference spectrum shows the element by element

difference of the normalized mass spectrum obtained from the labeled chromatogram and the normalized mass spectrum obtained from the unlabeled chromatogram.

2.3 Importing data

Data can be imported either in netCDF or MetaboliteDetector format. In case netCDF file format is used NTFD performs a deconvolution and compound detection step. However, it is highly recommended to preprocess the data with MetaboliteDetector.

2.4 Settings

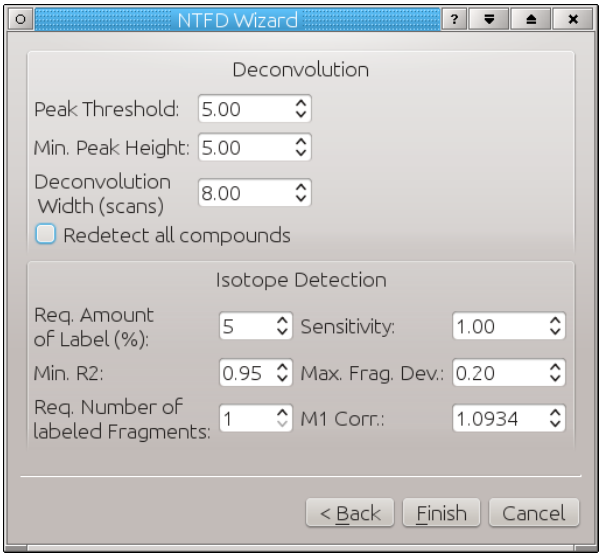

Figure 6: Settings window with deconvolution and isotope detection parameters.

2.4.1 Deconvolution

Peak Threshold The lower this value the higher the sensitivity

Min Peak Height Defines the minimal height (signal-to-noise units) of the peak maximum above the baseline that a compound must exceed to be detected. The lower this value the higher the sensitivity.

Deconvolution Width (scans) Defines the filter width (number of scans), which is used to match single ion peaks to a compound. The lower this value the better coeluting compounds are separated

Redetect all compounds If this checkbox is selected, NTFD will redetect all compounds. **Attention**: All compounds detected previously will be overwritten

2.4.2 Isotope detection

Req. Amount of label (%) The minimum relative amount of labeled substance ($1-M_0$) that has to be present to show the compound in the result list.

Min R2 The minimum correlation coefficient (R^2) value to show an putatively labeled ion in the results.

Req. Number of labeled fragments How many labeled fragments have to be present to consider a compound as valid.

Sensitivity Derivative of the difference spectrum has to be less than this value to consider a fragment as labeled.

Max. Frag. Dev Maximum allowed value of $1 - \sum_{i=0} \mid M_i \mid$.

M1 Corr. Correction for the natural isotope abundance (%). Default value is 13C natural abundance.

2.5 Recommended Settings

2.5.1 Agilent 6890 GC/5975B MS

- Peak threshold: 2–10 (2: high sensitivity, 10: low sensitivity)
- Minimum peak height: 2–10 (2: high sensitivity, 10: low sensitivity)
- Deconvolution width (scans): 7–8

2.5.2 Jeol AccuTOF

- Peak threshold: 5–20 (5: high sensitivity, 20: low sensitivity)
- Minimum peak height: 5–20 (5: high sensitivity, 20: low sensitivity)
- Deconvolution width (scans): 1

2.5.3 Trace

- Peak threshold: 50–100 (50: high sensitivity, 100: low sensitivity)
- Minimum peak height: 5-20 (5: high sensitivity, 20: low sensitivity)
- Deconvolution width (scans): 7–8

2.5.4 LECO GC/TOF

- Peak threshold: 20–50 (20: high sensitivity, 50: low sensitivity)
- Minimum peak height: 20–50 (20: high sensitivity, 50: low sensitivity)
- Deconvolution width (scans): 40

3 Step-by-step tutorial

In this tutorial we will use a small sample dataset. The dataset contains six files of a metabolomics experiment using the A549 human cancer cell line. Files `g1_1.cdf`, `g1_2.cdf` and `g1_3.cdf` contain GC-MS measurement data from cells cultivated in a medium containing 50% U-$^{13}C_6$-glucose and 50% unlabeled glucose. Files `g2_1.cdf`, `g2_2.cdf` and `g2_3.cdf` contain the data from cells cultured in normal unlabeled glutamine medium. The samples were derivatized with *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) and measured on an Agilent 6890 GC/5975B MS system.

To get started, download the dataset (`ntfd_sample_data.zip`) from the NTFD Website, unpack it, start NTFD and follow the steps below.

3.1 Start program and load samples

NTFD can be started by selecting *All applications*→*Education*→*Science*→*NTFD* (KDE) or by typing `ntfd-app` on the command line. On Windows select NTFD from the Start menu. The NTFD wizard will start automatically asking you to provide the labeled chromatograms. Click the and browse to the folder containing the downloaded data. Choose "netCDF" from the filter menu and select the three labeled chromatograms (`g1_1.cdf`, `g1_2.cdf` and `g1_3.cdf`). Then click next and select the unlabeled data (`g2_1.cdf`, `g2_2.cdf` and `g2_3.cdf`). In case you chose a wrong chromatogram you can remove it by clicking .

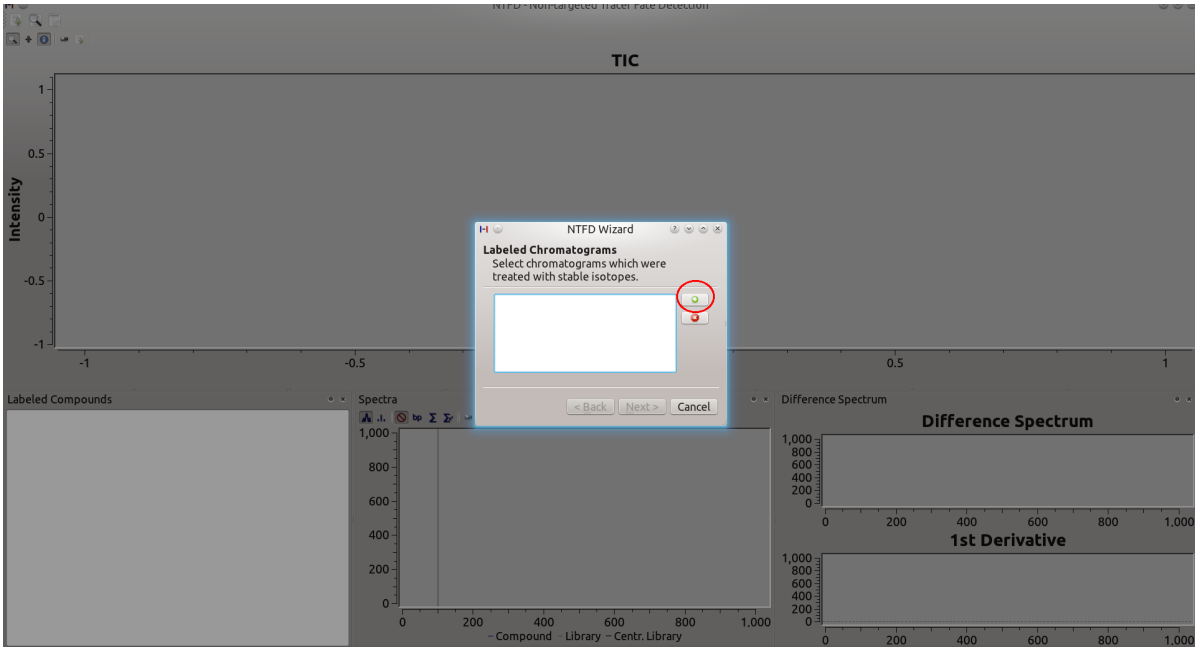

Figure 7: Step 1: NTFD wizard. File selection page.

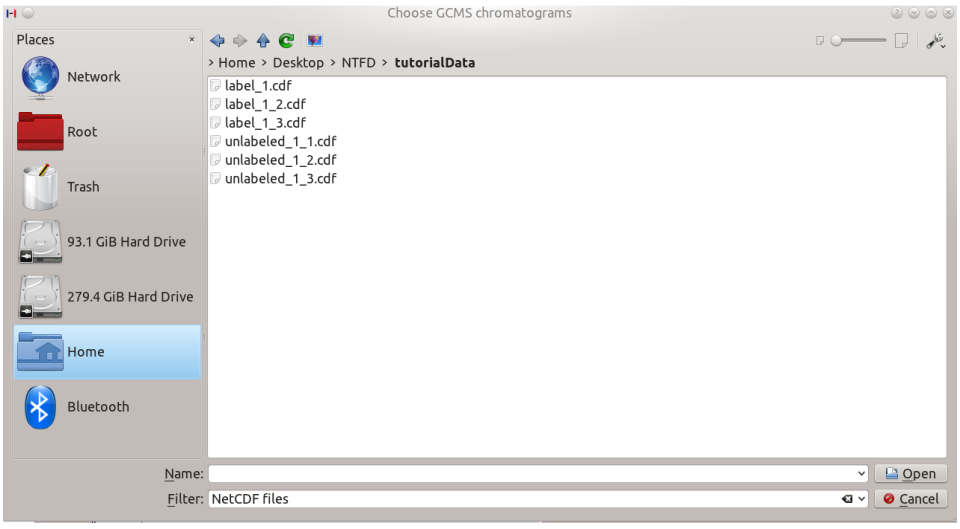

Figure 8: Step 2: Selection of labeled and unlabeled data.

3.2 Set the right settings

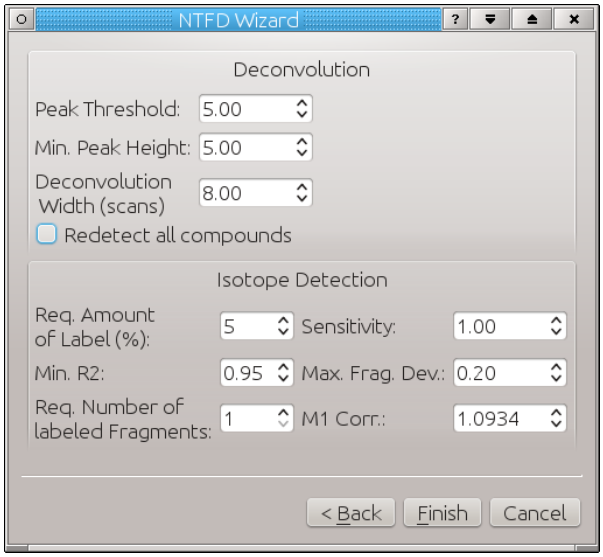

Figure 9: Settings window with deconvolution and isotope detection parameters.

The settings for deconvolution and label-detection are machine-specific. Enter the parameter settings shown in Fig. 9 and click finish. Detection of the labeled compounds can take some minutes depending on the parameter choice (sensitivity!) and your hardware.

3.3 Inspecting the results

The labeled compound list now shows all detected labeled compounds, their average retention time in the labeled and unlabeled chromatograms and the number of files in which the compound was detected (Spec. count). Clicking one compound shows the corresponding labeled and unlabeled mass spectra in the spectrum subwindow as well as their difference spectrum and its first derivative. Doubleclicking an entry expands the view and shows the m/z of the unlabeled fragment along with its relative mass isotopomer abundances (columns M+1, M+2, ...).

3.3.1 Compound identification

Until now compounds are presented as numbers only. They can be matched against a compound library (MetaboliteDetector format) to try to identify them. You can do this by clicking identify compounds in the mainwindow toolbar and selecting the small library `sample_library.lbr` in the sample data archive (Fig. 10). (You can examine the

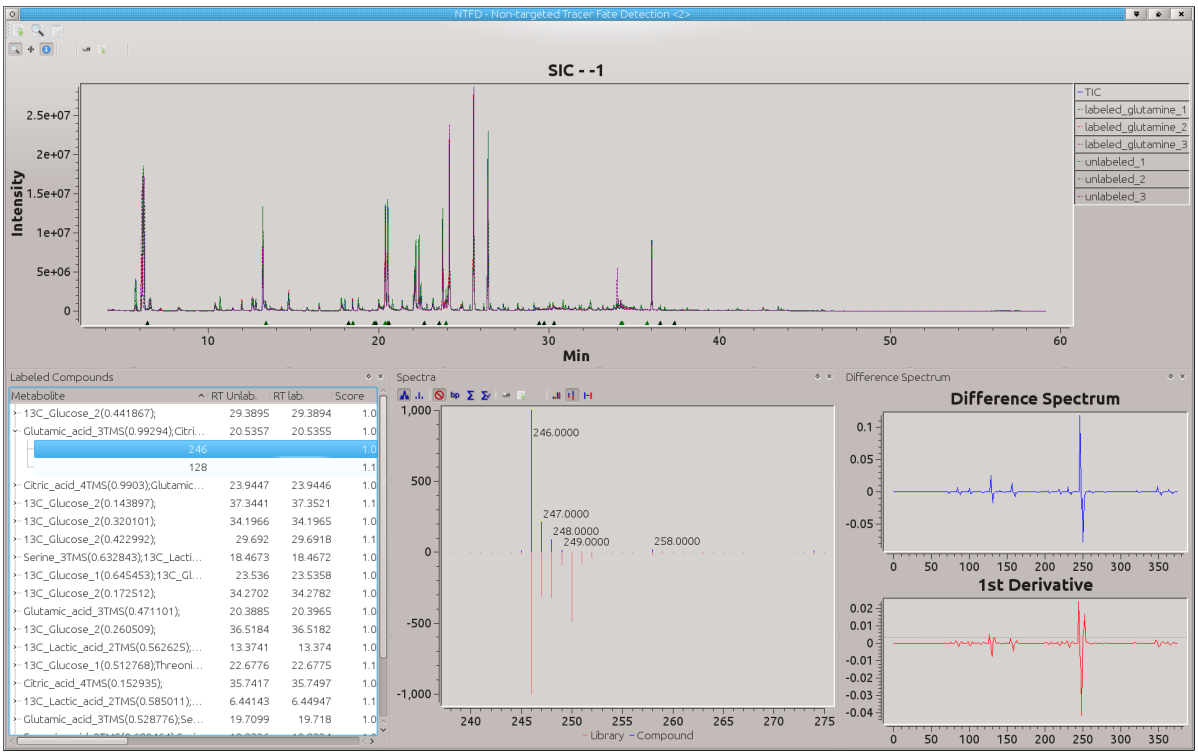

Figure 10: The processed sample data after compound identification.

library contents using the MetaboliteDetector software [1].) Different library compounds matching the measured spectra are presented sorted descending by their corresponding score which is shown in parentheses.

The score ($0 \leq score \leq 1$) is based on the normalized dot-product of the spectra. A score of 0 means no similarity, 1 is a perfect match. Generally scores > 0.95 are generally considered as good hits. The retention time information is currently not included in the score calculation, thus many false positive will show up. They can be spotted by their low identification score.

3.3.2 Looking at the sample data

In the sample data you will see many labeled compounds identified as Glucose. But there are only two rows with a decent identification score (i.e > 0.9). The two compounds are two different glucose derivatives. The different ions have different MIDs because they contain different numbers of carbon atoms. The ions with m/z 435 and 361 contain the full carbon backbone. The M+6 abundance roughly matches the 50% of U-$^{13}C_6$-glucose tracer contained in the media. The other fragments lost some of the backbone carbon.

Other good hits are fumaric acid and lactic acid. For fumaric acid only one small

labeled fragment was detected. To detect more fragments rerun NTFD with a lower sensitivity value. The dashed grey line in the 1st derivative plot represents the sensitivity threshold. For lactic acid two label-containing ions comprising the whole C_3 backbone were detected, m/z 219 and 203.

Many other spectra were found to contain tracer-derived ^{13}C but could not be identified using this small sample library. They can possibly be identified using bigger compound libraries or might be interesting candidates for structure elucidation.

3.3.3 Data export

To use the mass isotopomer data for further calculations it can be exported as tab-separated values. Choose from the mainwindow toolbar and enter a file name for the exported data. This data can now be viewed in spreadsheet applications and imported to statistics tools (Fig. 11). The data is presented in the same manner as in the NTFD result table. Additionally, the tsv file includes the 95% confidence intervals for the relative mass isotopomer abundances (columns CI+0, CI+1, CI+2, ...).

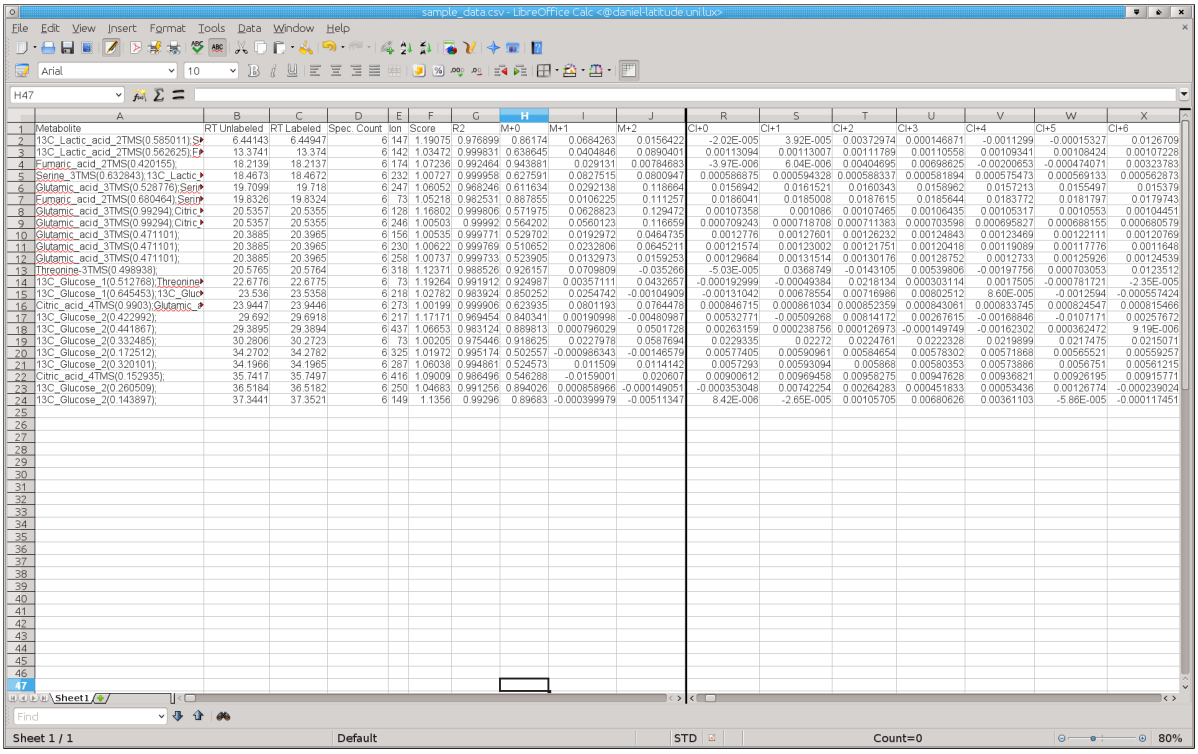

Figure 11: The data exported from NTFD can be imported into common spreadsheet applications.

The MID data can be used as a basis for metabolic flux analysis or other analyses.

4 References

References

[1] K. Hiller, J. Hangebrauk, C. Jäger, J. Spura, K. Schreiber, and D. Schomburg. Metabolitedetector: comprehensive analysis tool for targeted and nontargeted gc/ms based metabolome analysis. *Anal Chem*, 81(9):3429–3439, May 2009.

[2] K. Hiller, C. M. Metallo, J. K. Kelleher, and G. Stephanopoulos. Nontargeted elucidation of metabolic pathways using stable-isotope tracers and mass spectrometry. *Anal Chem*, 82(15):6621–6628, Aug 2010.