MIA is an application for non-targeted stable isotope labeling analysis. This document provides a basic overview over MIA 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), the underlying algorithms and important settings. The second part is a step-by-step tutorial that demonstrates the analysis of a sample dataset.

Additional information and the most recent version of MIA can be found at http://massisotopolomeanalyzer.lu/.

Daniel Weindl*
2014–2016
1 Introduction

MIA is a GUI-based software package for the non-targeted analysis of stable isotope labeling data. The detection and quantification of stable isotope labeling is based on the NTFD algorithm [3, 9]. MIA will detect all compounds derived from a stable isotope labeled tracer in a set of GC-MS chromatograms, visualize their corresponding mass isotopomer distributions (MIDs) and allow for further data analyses. MIA is implemented in C++ and is based on Qt5\(^1\), MetaboliteDetector [2], NTFD [3] and GraphViz\(^2\) [1]. Source code and binaries for Debian-based Linux distributions as well as for Windows systems are freely available at http://massisotopolomeanalyzer.lu/.

2 Background

Stable isotope labeling is widely used in metabolomics to infer metabolic fluxes or to assist chemical structure elucidation. After feeding a stable isotope labeled compound to an organism, the mass isotopomer distributions of metabolites are determined by the metabolic fluxes in the network. This is the basis for \(^{13}\text{C}\) metabolic flux analysis (\(^{13}\text{C}\)-MFA). To date, most mass isotopomer distribution analyses have been very targeted. MIA offers an easy-to-use interface for determination and analysis of stable isotope labeling GC-EI-MS data in a non-targeted manner.

MIA visualizes all detected MIDs, allows to cluster them by their similarity to find metabolically closely related compounds, offers filtering, compound identification and data export capabilities. Stable isotope labeling data from multiple datasets can be analyzed in parallel. The combined analysis of datasets from different tracers can be helpful for compound identification or pathway contextualization. The analysis of mass isotopomer distributions after feeding the same tracer under different experimental conditions can be used to globally analyze alterations in metabolic fluxes [7].

In [7], we show how MIA was applied to analyze metabolism of hypoxic cancer cells and how non-targeted stable isotope labeling analysis helps provide novel biological insights.

2.1 Non-targeted tracer fate detection (NTFD)

The detection of stable isotope labeling in MIA is based on the previously published NTFD algorithm [3, 4, 9]. For NTFD, an organism is grown on standard medium and in

\(^1\)http://qt-project.org/
\(^2\)http://www.graphviz.org/
parallel in the same medium with one or more components fully or partially replaced by their stable isotope enriched analogue. Metabolites from both cultures are extracted and subjected to GC-MS analysis. For each analyte, the mass spectra from the isotopically enriched and non-enriched sample are matched. From the difference of these mass spectra isotopically enriched fragments can be detected. For each of these fragments the mass isotopomer distribution can be calculated.

2.2 MID-based compound networks

A central feature of MIA is the visualization of labeled compound as MID-similarity-based networks. Metabolically closely related compounds show very similar MIDs. Thus, MIA analyzes the similarity of MIDs and connects compounds with highly similar MIDs which are a sign of proximity within the metabolic network. MIDs are compared in a pair-wise manner. First, a Needleman-Wunsch-Alignment is performed on the MID vectors to account for gains or losses of isotopically enriched groups. Then, the distance of these aligned vectors is calculated and the user can select a similarity threshold for which metabolites should be considered as connected.

The following distance measures are implemented in MIA:

- Euclidean distance
  \[ d(p, q) = \sqrt{\sum_{i=0}^{n} (p_i - q_i)^2} \]

- Canberra distance
  \[ d(p, q) = \sum_{i=0}^{n} \frac{|p_i - q_i|}{|p_i| + |q_i|} \]

- Manhattan distance
  \[ d(p, q) = \sum_{i=0}^{n} |p_i - q_i| \]

where \( p \) and \( q \) are the aligned MID vectors in \( \mathbb{R}^n \) of the first and second analyte respectively. The MID distances can be normalized to the length of the MID vectors.

2.3 MID variation filtering

When comparing stable isotope labeling after feeding the same tracer under different experimental conditions, it is interesting to detect changes in mass isotopomer distributions since they are indicative of changes in metabolic fluxes [8, 7]. To detect compounds with
varying labeling patterns, MIA allows to filter the results by variation in their MIDs. Therefore the maximum standard deviation in relative mass isotopomer abundance is determined for every compound and the user can hide compounds with low variation score (3.2.8).

\[
\text{variation score} = \max \sigma_i \quad \sigma_i = \sqrt{\frac{1}{n} \cdot \sum_{j=0}^{n} (\overline{p}_i - p_{i,j})^2}
\]

where \( p_{i,j} \) is the relative abundance of the M + j isotopologue of the given compound in the \( i \)-th dataset. The MIDs of unlabeled compounds are not considered. 

3 User manual

3.1 Installation

Installation packages for MIA can be downloaded from [http://massisotopolomeanalyzer.lu/](http://massisotopolomeanalyzer.lu/) in the .deb or .exe format. Alternatively, you can download a plain archive with the binaries.

3.1.1 Linux

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

\[\text{sudo dpkg-i filename.deb}\]

for Debian based systems

where filename is the name of the deb installation package (e.g. mia-1.0-Linux.deb).

3.1.2 Windows

Unpack the MIA-Windows archive and run mia-gui.exe

3.2 Graphical user interface (GUI)

The graphical user interface is organized in one main window containing three subwindows which are explained below. Several application settings can be changed in the configuration dialog (3.2.10).
3.2.1 Main window

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

**Main toolbar**

- **Import data** Opens the data import panel (3.2.12) to import GC-MS data from the netCDF format or perform peak detection on chromatograms in the MetaboliteDetector [2] format.

- **Open XML** Load dataset from MIA XML configuration file.

- **Save binary data** Save current dataset to MIA binary data file.

- **Export MIDs to CSV** Save MIDs of all detected labeled fragments to comma separated values (CSV) file. The exported data can be opened with common spreadsheet programs like OpenOffice Calc or Microsoft Excel.

\[\text{Compounds which are only labeled under specific conditions are also of interest. However, they need to be checked for false positives.}\]
Library search  Try to identify the detected compounds (see 3.2.4) based on a MetaboliteDetector reference library. Matching is performed on spectrum similarity and retention index available (3.2.9).

About  Show MIA version information.

Application settings  Shows the settings dialog (see 3.2.10).

3.2.2  Graph view

The central component of the MIA GUI is the graph view. Here you can find a lot of information on all labelled compounds detected in the active datasets. The layout and the compounds shown depend on the settings in the experiment panel (3.2.5), the graph panel (3.2.8) and the settings dialog (3.2.10). For every compound the MID of the selected labeled fragment (usually the largest) is shown for each active dataset (3). The background color of the ellipse reflects the variation in relative mass isotopomer abundance, ranging from dark blue (low) to high (white). The MID plots for each dataset set show the relative mass isotopomer abundances as bars with 95% confidence intervals and numbers, the compound name (after identification - see 3.2.9), the selected $m/z$, and some quality measures. The background color reflects the amount of isotopic enrichment $(1 - M_0)^4$, ranging from dark gray (unlabeled) to very light gray (fully labeled). As quality measures, there are the coefficient of determination $R^2$ (“R2”) and the sum of absolute mass isotopomer abundances (“S”). Both these values should ideally be equal to 1 [9]. Thresholds for both can be defined in the label detection parameters in the experiment wizard (3.2.6).

Double click the oval to pop-up the detailed compound view (3.2.3) (you can open multiple windows) or hold the mouse for a tooltip with additional information.

Navigation  You can zoom in and out of the graph view by holding [CTRL] and using the mouse wheel. You can reset the zoom by clicking \(\mathcal{Q}\) in the toolbar. To navigate through the graph view, you can use the scroll bars, or the mouse wheel or scroll keys, or move the mouse while holding the middle button.

Toolbar actions

\(\mathcal{Q}\)  Reset zoom  Reset zoom to fit all plots into the view.

\[^4\]The value of $1 - M_0$ is used, since the actual isotopic enrichment $1/n \cdot \sum i \cdot M_i$ cannot be determined because the number of possibly labeled atoms $n$ is not known before compound identification and knowledge on fragmentation.
Figure 2: The graph view.

Export image Export the current graph as scalable vector graphics (SVG). The exported file can easily be edited with most vector graphics applications like e.g. Inkscape\(^5\). Individual plots can easily be copied and modified from there to include them in other documents or presentations.

3.2.3 Compound detail view

Double-clicking a node in the graph pops up the compound details view (Fig. 4). There, you can find MID plots for all isotopically enriched fragments in all active datasets and

\(^5\)http://www.inkscape.org/
some additional information.

Below the name, you can find the datasets in which the given compound was found to show isotopic enrichment along with its retention time (RT) and retention index (RI). The array of plots shows the MIDs of all labeled fragments (in rows) in all active datasets (in columns). The plot details are explained above (3.2.2). Double-clicking a plot will copy the corresponding mass MID to the system clipboard.

3.2.4 Compound panel

The compound panel (Fig. 5) is displayed in left part of the MIA main window. It lists all compounds which show isotopic enrichment in any of the currently active datasets. MIDs of all labeled fragments can be viewed after clicking the expand icon in the tree view. Click on the compound name to center the graph view (2) on this compound.

The toolbar provides the following functionality:

🔍 Expand Expand the tree view and see the following information:

<table>
<thead>
<tr>
<th>+ Compound name</th>
<th># datasets</th>
<th># ions</th>
<th>$R^2$</th>
<th>abundance</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>⊂ Dataset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊂ Labelled fragment $m/z$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊂ Mass isotopomer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This information can be exported as CSV from the main toolbar (CSV - 3.2.1).

🔍 Collapse Collapse the tree view to only show the list of labelled compounds.
Figure 4: The compound details window showing MID plots for all isotopically enriched fragments in all active datasets and some additional information.

### 3.2.5 Dataset panel

This panel shows the currently loaded datasets and allows you to add or remove new datasets (Fig. 6). The table shows the name of your dataset and the number of isotopically enriched compounds and non-enriched compounds. With the checkbox in the first column you can select which datasets should be used for the network generation (2.2 and 3.2.8).

To add a new dataset hit “Add” and follow the experiment wizard (3.2.6). To remove a dataset, select it in the table and hit “Remove”. The settings for label detection and compound identification can be edited by double clicking the corresponding dataset in the table (3.2.6).

### 3.2.6 Add and Edit datasets

New datasets are added or existing datasets are modified via the experiment wizard which can be opened by clicking “Add” or double clicking a dataset in the experiment panel (3.2.5). The first page asks for a name and the data files for the dataset (Fig. 3.2.6).
Name  An arbitrary name for the dataset.

(Un)Labeled chromatograms  Data files from the measurements of the (un)labelled metabolite extract (2.1). If multiple data files are added here, confidence intervals can be calculated for the MIDs [3]. Data from multiple independent experiments should added as separate dataset, not as multiple chromatograms for the same dataset.

On the second wizard page, a number of parameters can be set for the detection of isotopic enrichment, compound identification, and a few more things.\(^6\)

Detection of isotopic enrichment:\(^7\)

Maximum fragment deviation  The maximum allowed value for \(1 - \sum |M_i|\) (important quality filter).

Minimum number of labeled fragments  Compounds with a number of labeled fragments lower than this number will be ignored.

Required amount of isotopic enrichment  Minimum value for \(1 - M_0\).

Minimum \(R^2\)  Minimum allowed coefficient of determination \(R^2\) for each fragment.

---

\(^6\)The default settings are a good starting point for most applications, but should be adjusted for optimal results.

\(^7\)See also NTFD documentation available at [http://ntfd.mit.edu/](http://ntfd.mit.edu/) and [9]
Figure 6: List of currently active datasets.

**Minimum M0 abd.** The minimum relative abundance of the M+0 mass isotopomer.

**Ignore compounds with M\_n \mid with n \geq \ldots** The upper limit for MID size.

**Compound identification:**

**RI tolerance** Changes the influence of the retention index (RI) on the compound identification score.

**Matching score cutoff** Defines the threshold for the spectrum matching score above which a library compound is considered as a hit.

**Show top n hits** Defines how many compound identification hits should be kept as the label.

**Others:**

**Gap penalty for Needleman-Wunsch-Scoring** Gap penalty used for the Needleman-Wunsch alignment of MIDs for the distance calculation (2.2).

**Distance cutoff graph edges** Distance cut-off for MID similarity analysis (2.2). This value should not be modified here, but in the graph settings panel (3.2.8).
3.2.7 Save XML

To avoid repeatedly selecting the data files and adjusting the settings you can save the settings in an XML file which can later be reloaded by clicking “Open XML” in the main toolbar.

Note: MIA saves absolute paths to the data in the .xml files. If the data was moved, the paths can be adjusted in the .xml file using a standard text editor.

3.2.8 Graph panel

The graph panel (Fig. 8) contains the settings for the generation of the compound networks (2.2).

**Layout engine**  The GraphViz engine to layout the graph. See GraphViz documentation for more information. “dot” is the default and recommended.

**Distance metric**  Metric used to calculate MID distance (2.2).

**Distance normalization**  Optional normalization factor for MID distances.
Figure 8: Graph panel to setup the compound network to connect compounds by MID similarity (2.2).
Distance cutoff MIDs with distances below the given threshold are considered as connected in the compound graph.

Minimum experiments The minimum number of datasets in which the compound has to be detected as isotopically enriched to be considered for the compound graph. Select “Hide others” to hide all nodes detected as labeled in a lower number of datasets.

Variation cutoff The minimum value for the variation score (2.3) for each node to be considered for the compound graph. Select “Hide others” to hide all nodes with a lower variation score.

3.2.9 Compound identification

The detected isotopically enriched compound can be matched against a mass spectrum library for compound identification by selecting “Library search” from the main toolbar. Currently, only MetaboliteDetector [2] libraries are supported.8

MIA uses a spectrum and RI / RT based algorithm for spectrum matching. The RI tolerance and identification score cutoff can be defined in the experiment wizard (3.2.6).

3.2.10 Settings

The settings dialog (Fig. 3.2.10) can be opened by clicking in the main toolbar (3.2.1). More graph options and certain default paths can be set here (Fig. 3.2.10).

Show unconnected nodes Toggle whether nodes which are not connected to any others nodes should be shown or not (see also 3.2.8 and 2.2).

Use largest common ion If a compound was detected in multiple datasets, the heaviest labeled fragment common to all datasets can be used in the plot and for distance calculation, or the heaviest fragment for each dataset independently.

MID plot layout MID plots for a compound labeled in multiple experiments can be laid out vertically, horizontally or in a quadratic grid.

8If you do not have a compound library in the MetaboliteDetector format available already, the Golm Metabolome Database (GMD) [5] provides a good start. Download their mass spectral database in the MSL format from http://gmd.mpimp-golm.mpg.de/download/ and create a library file from it using MetaboliteDetector (http://metabolitedetector.tu-bs.de/) by selecting File → Import → Import MSL library. The resulting .lbr file can be used with MIA.
Figure 9: The configuration page for paths and miscellaneous settings.

**Distance calculation** Defines on which data the MID distance calculation is performed. The full MID vector can be used (“Include”) or the M+0 can be omitted (“Exclude M0”). If M+0 is omitted, the remaining vector can be divided by the highest value (“normalize to basepeak”) or the sum (“normalize to sum”).

### 3.2.11 Data export

An important feature of MIA are its data export capabilities. The following data can be exported:

- **MIDs of all compounds detected as isotopically enriched** By selecting "Export MIDs to CSV" from the main toolbar
- **MID plots of all compounds and the MID similarity-based network** By selecting "Export image" in the graph view toolbar (3.2.2)
- **MID plots for all fragments of each compound** By selecting "Export image" in the Compound detail view (3.2.3)
- **The settings used for the analysis** By selecting "Save XML" in the dataset panel (3.2.5)

### 3.2.12 Data import

MIA requires data in the MetaboliteDetector format. GC-MS data can be imported from the netCDF format via the data import dialog (Fig. 10) which can be accessed via
"Import data" in the main toolbar. This dialog can also be used to re-perform peak picking and spectrum deconvolution on existing MetaboliteDetector data.

Note: For best control over deconvolution, these steps should be performed in MetaboliteDetector where they can be inspected immediately. Low quality spectra should already be filtered at this stage to increase specificity and speed of the detection of isotopic enrichment.

Figure 10: Data import window
4 Step-by-step tutorial

On the MIA website, there is a sample data set available for download. In the following, we will demonstrate the first steps in MIA to explore this stable isotope labeling data.

The sample data set contains GC-MS measurements of trimethylsilylated metabolite extracts from A549 cells incubated under 21% or 1% oxygen atmosphere ("N/"H") in the presence of [U-\textsuperscript{13}C]glucose ("U\textsuperscript{13}CGlc"), [U-\textsuperscript{13}C]glutamine ("U\textsuperscript{13}CGln") or only unlabeled substrates ("" ). The data is supplied ready-to-use in the MetaboliteDetector format (.bin/.idx/.cmp).

4.1 Start program and load data

Download and unpack the sample data and run MIA. For a quick start, the file sample-Data.xml contains all parameters for the given data sets and can be loaded by selecting "Open XML" from the main toolbar.

Note: sampleData.xml must be located in same folder as the data files.

After selecting the .xml file, compound detection will start automatically. Depending on your system, this might take up to a few minutes. Once, compound detection is finished, you will see all compounds which were detected as isotopically enriched (Fig. 11).

Figure 11: MIA after loading the sample dataset.
4.2 Inspecting the results

In the dataset panel, you can see the different experimental conditions and how many compounds were detected as isotopically enriched. In the center, you can see every detected enriched compound as an ellipse with plotted MIDs for each experimental condition with the same color coding as on the list on the right (see 3.2.2). The compound panel on the left shows a list of all compounds which were detected in any of the datasets. For now, these compounds are all named by their retention index (RI).

→ Scroll around and zoom in to inspect the MID plots (3.2.2). Double click them to get more detailed information. When you double click on an entry in the dataset list, you can see which files are associated and can change parameters for the detection of isotopic enrichment (3.2.6).

4.2.1 Compound identification

To find out which compounds were detected as isotopically enriched, you can match their spectra against a library with known mass spectra (3.2.9). To this end, select “Library search” from the main toolbar, locate the file sample-lib-TMS.lbr which was part of the sample data archive. Select “No” in the next dialog. Click on heading of the first column of the compound list to order compounds alphabetically. The result should look similar to Fig. 12.

4.2.2 MID variation

You can use MIA to detect those compounds with changing MIDs, which are indicative of changing fluxes in response to lower oxygen levels. To obtain meaningful results, only data from the same tracer should be used. Let’s analyze isotopic enrichment from \([U-^{13}C]glutamine\). Therefore, first remove the \([U-^{13}C]glucose\) labeling datasets by selecting them in the data set list and clicking “remove”. For the MID variation analysis (2.3), ideally the MIDs from the same mass spectrometric fragment is analyzed. To ensure this, access the configuration dialog from the main toolbar and select Misc → Use largest common ion.

Check “Hide others” under “Variation cutoff” and move the slider slowly to the right. Some ellipses will start to disappear. Stop, when there are only 5 compounds left. Use the sample library as above to try to identify these compounds (Fig. 13).

Malic acid and aspartic acid show a strong change in MIDs in response to lower oxygen availability. The strong increase in relative M+3 abundance and the strong decrease in M+4 abundance are indicative of an inversion of isocitrate dehydrogenase
flux directionality, an important feature of cancer hypoxic cancer cells [10, 6, 7].

4.2.3 MID similarity

Three other compounds of the five selected in the previous step are still unidentified. We try to gain information on their metabolic origin by comparing their MIDs to those of other compounds. Let’s try that for the compound RI 1782.98.

In the configuration dialog, uncheck Misc $\rightarrow$ Use largest common ion. In the Graph options, uncheck “Hide others” under “Variation cutoff” and move slider back to 0. Now move the “Distance cutoff” slider slowly to the right, to about 1%. When the graph is plotted, click on “RI 1782.98” in the compound list to center the view on it. The result will look similar to Fig. 14.

From the network you can see that our compound of interest RI 1782.98 is has a extremely similar MID as compound RI 1705.96. This is a hint, that both compounds are part of the same linear metabolic pathway, or that these are two derivatives arising during sample preparation from the same native metabolite. Furthermore, aspartic acid and malic acid show high MID similarity at normoxia, and thus, seem to be closely related.

And indeed, a more comprehensive reference library will confirm these observations: RI 1782.98 and 1705.96 are the 3TMS and 2TMS derivative of N-acetylaspartic acid.
Figure 13: Hypoxia-induced changes after [U-^{13}C]glutamine labeling.

Figure 14: MID-neighbors of RI 1782.98.
respectively. N-acetylaspartic acid is derived from aspartic acid which is in turn derived from malic acid, explaining the high observed MID similarity.

4.2.4 Exporting the data

Export the MIDs of all compound as csv file, by selecting “Export MIDs to CSV” from the main toolbar. Export the MID plots currently shown in the graph view by selecting “Export image” in the graph view toolbar (3.2.2). Inspect the exported data in a spreadsheet application (.csv) and a vector graphics editor (.svg).

References


