Metabolite identification

Michael Witting
Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München
&
Chair of Analytical Foodchemistry, TU München
Outline

A path to metabolite identification

- Introduction
  - Analytical Technologies
  - Terminology in metabolite identification
- Metabolite identification
  - GC-MS MetID
    - EI spectral deconvolution
    - GC-APCI and exact mass
  - LC-MS MetID
    - Metabolite / Ion annotation
    - Formula calculation
    - Utilization of isotopes for formula calculation
    - MS/MS and RT
- New approaches for MetID
  - Ion mobility
  - in silico approaches
- de novo identification
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • *in silico* approaches

• *de novo* identification
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
    • Terminology in metabolite identification
• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT
• New approaches for MetID
  • Ion mobility
  • in silico approaches
• de novo identification
Introduction

Analytical Technologies

- Metabolomics utilizes different analytical chemistry methods for analysis of metabolites
- These methods include
  - **Mass Spectrometry (MS)** without or with prior metabolite separation
    - **Gas Chromatography (GC)**
    - **Liquid Chromatography (LC)**
    - **Capillary Electrophoresis (CE)**
  - **Nuclear Magnetic Resonance (NMR)**
- Each method has its specific advantages and disadvantages
- **No method can cover all metabolites!**
- Certain overlap between the different methods exists
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • in silico approaches

• de novo identification
Introduction

Terminology

• Following cases have to be consider once it comes to metabolite identification

  • *Known metabolites* = known structure, present in matrix, detected by method of choice

  • *Known unknown metabolites* = know structure, shown to be not present in in matrix so far, detected by method of choice, e.g. comparison of MS/MS spectra

  • *Unknown metabolites* = *de novo* identification of truly novel compounds
Introduction

Terminology

• Not every single metabolite gives a single peak
• GC-MS
  • Different derivatization products
  • In source fragments (EI is a hard ionization technique)
• Although ESI is a soft ionization technique many signals are generated in LC-MS, SFC-MS and CE-MS for a single metabolite
  • Isotopes, adducts and multimers
  • In source fragments
  • Different charge states
• NMR
  • Different shifts and coupling
• It is important to differentiate between
  • Annotation = potential metabolites, e.g. based on comparison with exact masses
  • Identification = proven metabolites
**Introduction**

**Terminology – GC-MS multiple peaks**

**A**

- Ionization Region
- Anode
- Accelerating Slits
- Sample
- Ions to Analyzer
- Repeller
- Filament

\[
M + e^- \rightarrow M^+ + 2e^- \\
\rightarrow M_1^+ + M_2^+
\]

\[
M_3^+ + M_5^+ + M_7^+ \\
+ + + \\
M_4^+ + M_6^+ + M_8^+
\]

**B**

- Ionization Region
- Anode
- Accelerating Slits
- Sample
- Ions to Analyzer
- Repeller
- Filament
- Reagent Gas

\[
CH_4 + e^- \rightarrow CH_4^+ + 2e^- \\
CH_4 + CH_4^+ \rightarrow CH_5^+ + CH_\cdot_3 \\
CH_4 + CH_3^+ \rightarrow C_2H_5^+ + H_2 \\
M + CH_5^+ \rightarrow CH_4 + [M + H]^+ \\
AH + CH_3^+ \rightarrow CH_4 + A^+
\]
Introduction

Terminology – LC-MS multiple peaks
Introduction

Terminology – LC-MS multiple peaks

<table>
<thead>
<tr>
<th>M+3H</th>
<th>M/3 + 1.007276</th>
<th>3+</th>
<th>0.33</th>
<th>1.007276</th>
</tr>
</thead>
<tbody>
<tr>
<td>M+2H+Na</td>
<td>M/3 + 8.334590</td>
<td>3+</td>
<td>0.33</td>
<td>8.334590</td>
</tr>
<tr>
<td>M+H+2Na</td>
<td>M/3 + 15.766190</td>
<td>3+</td>
<td>0.33</td>
<td>15.766190</td>
</tr>
<tr>
<td>M+3Na</td>
<td>M/3 + 22.989218</td>
<td>3+</td>
<td>0.33</td>
<td>22.989218</td>
</tr>
<tr>
<td>M+2H</td>
<td>M/2 + 1.007276</td>
<td>2+</td>
<td>0.50</td>
<td>1.007276</td>
</tr>
<tr>
<td>M+H+NH4</td>
<td>M/2 + 9.520550</td>
<td>2+</td>
<td>0.50</td>
<td>9.520550</td>
</tr>
<tr>
<td>M+H+Na</td>
<td>M/2 + 11.998247</td>
<td>2+</td>
<td>0.50</td>
<td>11.998247</td>
</tr>
<tr>
<td>M+H+K</td>
<td>M/2 + 19.985217</td>
<td>2+</td>
<td>0.50</td>
<td>19.985217</td>
</tr>
<tr>
<td>M+ACN+2H</td>
<td>M/2 + 21.520550</td>
<td>2+</td>
<td>0.50</td>
<td>21.520550</td>
</tr>
<tr>
<td>M+2Na</td>
<td>M/2 + 22.989218</td>
<td>2+</td>
<td>0.50</td>
<td>22.989218</td>
</tr>
<tr>
<td>M+2ACN+2H</td>
<td>M/2 + 42.033823</td>
<td>2+</td>
<td>0.50</td>
<td>42.033823</td>
</tr>
<tr>
<td>M+3ACN+2H</td>
<td>M/2 + 62.547097</td>
<td>2+</td>
<td>0.50</td>
<td>62.547097</td>
</tr>
<tr>
<td>M+H</td>
<td>M + 1.007276</td>
<td>1+</td>
<td>1.00</td>
<td>1.007276</td>
</tr>
<tr>
<td>M+NH4</td>
<td>M + 18.033823</td>
<td>1+</td>
<td>1.00</td>
<td>18.033823</td>
</tr>
<tr>
<td>M+Na</td>
<td>M + 22.989218</td>
<td>1+</td>
<td>1.00</td>
<td>22.989218</td>
</tr>
<tr>
<td>M+CH3OH+H</td>
<td>M + 33.033489</td>
<td>1+</td>
<td>1.00</td>
<td>33.033489</td>
</tr>
<tr>
<td>M+K</td>
<td>M + 38.963158</td>
<td>1+</td>
<td>1.00</td>
<td>38.963158</td>
</tr>
<tr>
<td>M+ACN+H</td>
<td>M + 42.033823</td>
<td>1+</td>
<td>1.00</td>
<td>42.033823</td>
</tr>
<tr>
<td>M+2Na-H</td>
<td>M + 44.971160</td>
<td>1+</td>
<td>1.00</td>
<td>44.971160</td>
</tr>
<tr>
<td>M+IsoProp-H</td>
<td>M + 61.06534</td>
<td>1+</td>
<td>1.00</td>
<td>61.06534</td>
</tr>
<tr>
<td>M+ACN+Na</td>
<td>M + 64.015765</td>
<td>1+</td>
<td>1.00</td>
<td>64.015765</td>
</tr>
<tr>
<td>M+2K-H</td>
<td>M + 76.919040</td>
<td>1+</td>
<td>1.00</td>
<td>76.919040</td>
</tr>
<tr>
<td>M+DMSO+H</td>
<td>M + 79.02122</td>
<td>1+</td>
<td>1.00</td>
<td>79.02122</td>
</tr>
<tr>
<td>M+2ACN+H</td>
<td>M + 83.060370</td>
<td>1+</td>
<td>1.00</td>
<td>83.060370</td>
</tr>
<tr>
<td>M+IsoProp+Na+H</td>
<td>M + 84.05511</td>
<td>1+</td>
<td>1.00</td>
<td>84.05511</td>
</tr>
<tr>
<td>2M+H</td>
<td>2M + 1.007276</td>
<td>1+</td>
<td>2.00</td>
<td>1.007276</td>
</tr>
<tr>
<td>2M+NH4</td>
<td>2M + 18.033823</td>
<td>1+</td>
<td>2.00</td>
<td>18.033823</td>
</tr>
<tr>
<td>2M+Na</td>
<td>2M + 22.989218</td>
<td>1+</td>
<td>2.00</td>
<td>22.989218</td>
</tr>
<tr>
<td>2M+K</td>
<td>2M + 38.963158</td>
<td>1+</td>
<td>2.00</td>
<td>38.963158</td>
</tr>
<tr>
<td>2M+ACN+H</td>
<td>2M + 42.033823</td>
<td>1+</td>
<td>2.00</td>
<td>42.033823</td>
</tr>
<tr>
<td>2M+ACN+Na</td>
<td>2M + 64.015765</td>
<td>1+</td>
<td>2.00</td>
<td>64.015765</td>
</tr>
</tbody>
</table>
Introduction

Terminology – LC-MS multiple peaks
Introduction

Terminology – NMR multiple peaks

<table>
<thead>
<tr>
<th>Chemical shift</th>
<th>Net intensity</th>
<th>Multiplet information</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.87 ppm</td>
<td>2</td>
<td>s</td>
</tr>
<tr>
<td>7.47 ppm</td>
<td>2</td>
<td>t</td>
</tr>
<tr>
<td>7.53 ppm</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>7.80 ppm</td>
<td>2</td>
<td>q</td>
</tr>
<tr>
<td>8.36 ppm</td>
<td>1</td>
<td>s</td>
</tr>
<tr>
<td>12.79 ppm</td>
<td>1</td>
<td>s</td>
</tr>
</tbody>
</table>
### MSI levels of identification

<table>
<thead>
<tr>
<th>Level</th>
<th>Confidence of identity</th>
<th>Level of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Confidently identified compounds</td>
<td>Comparison of two or more orthogonal properties with an authentic standard analyzed under identical analytical conditions</td>
</tr>
<tr>
<td>2</td>
<td>Putatively annotated compounds</td>
<td>Based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries, without reference to authentic chemical standards</td>
</tr>
<tr>
<td>3</td>
<td>Putatively annotated compound classes</td>
<td>Based upon characteristic physicochemical properties of a chemical class of compounds or by spectra similarity to known compounds of chemical class</td>
</tr>
<tr>
<td>4</td>
<td>Unknown compounds</td>
<td>Although unidentified and unclassified, these metabolites can still be differentiated and quantified based upon spectra data</td>
</tr>
</tbody>
</table>
Introduction

MSI levels of identification

• Sumner et al. suggested a scoring system for metabolite identification

• Based on different scores for different analytical techniques you “quantitatively” report your confidence

• The sum of individual scores would give the final score

• If data is compared to an authentic standard the score is multiplied with 2, if compared to a public database the score is multiplied with 1.5

• Example 1: UPLC-MS with accurate mass and comparison with an authentic standard

• \((1.5 + 1.0) \times 2 = 5.0\)

• Example 2: GC-MS with high resolution RI, nominal mass and match to NIST library

• \((1.5 + 0.5 + 1.5) \times 1.5 = 5.25\)
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • \textit{in silico} approaches

• \textit{de novo} identification
Outline

A path to metabolite identification

- Introduction
  - Analytical Technologies
  - Terminology in metabolite identification

- Metabolite identification
  - GC-MS MetID
    - El spectral deconvolution
    - GC-APCI and exact mass
  - LC-MS MetID
    - Metabolite / Ion annotation
    - Formula calculation
    - Utilization of isotopes for formula calculation
    - MS/MS and RT

- New approaches for MetID
  - Ion mobility
  - in silico approaches

- de novo identification
Metabolite Identification

**GC-MS MetID – EI spectral deconvolution**

**A**

- Ionization Region
- Anode
- Accelerating Slits
- Sample
- Repeller
- Filament

\[ M + e^- \rightarrow M^+ \cdot + 2 e^- \]

\[ M^1_+ + M_2^- \]

\[ M^3_+ + M_5^+ + M_7^+ \]

\[ M_4 + M_6 + M_8 \]

**B**

- Ionization Region
- Anode
- Accelerating Slits
- Sample
- Repeller
- Filament
- Reagent Gas

\[ CH_4 + e^- \rightarrow CH_4^+ \cdot + 2 e^- \]

\[ CH_4 + CH_4^+ \rightarrow CH_5^+ + CH_3^- \]

\[ CH_4 + CH_3^+ \rightarrow C_2H_5^+ + H_2 \]

\[ M + CH_5^+ \rightarrow CH_4 + [M + H]^+ \]

\[ AH + CH_3^+ \rightarrow CH_4 + A^+ \]
Metabolite Identification

GC-MS MetID – EI spectral deconvolution

- Fragmentation-rich ionization methods like EI lead to mixed mass spectra
- Mixed spectra have to be deconvoluted into individual compound spectra
- AMDIS is a software capable of deconvoluting GC-MS runs
- Deconvoluted compounds and compound spectra can be searched in EI libraries, e.g. NIST, Golm Metabolome Database, Fiehn library, etc.
Outline

A path to metabolite identification

- Introduction
  - Analytical Technologies
  - Terminology in metabolite identification

- Metabolite identification
  - GC-MS MetID
    - EI spectral deconvolution
    - GC-APCI and exact mass
  - LC-MS MetID
    - Metabolite / Ion annotation
    - Formula calculation
    - Utilization of isotopes for formula calculation
    - MS/MS and RT

- New approaches for MetID
  - Ion mobility
  - In silico approaches
Metabolite Identification

*GC-MS MetID – GC-APCI and exact mass*

Analytics of LC unamenable compounds

Benzo(a)pyren

No hit in EI library search?

Libraries (NIST)
Metabolite Identification

*GC-MS MetID – GC-APCI and exact mass*

**GC-APCI-TOF/MS coupling**
- Soft ionization
- Molecular ion preserved
- High resolution data
- ID by SmartFormula
Metabolite Identification

**GC-MS MetID – GC-APCI and exact mass**

C\textsubscript{18}H\textsubscript{41}O\textsubscript{7}Si\textsubscript{4}

1.2 ppm

8.2 mSigma
Metabolite Identification

**GC-MS MetID – GC-APCI and exact mass**

- GC-MS spectral databases still contain only a limited number of APCI-QToF spectra
- Ruttkies et al. developed a workflow based on in silico derivatization of metabolites and in silico fragmentation to identify potential candidate structures

Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • in silico approaches
  • de novo identification
Metabolite Identification
LC-MS MetID - Metabolite / Ion annotation

MassTRIX: Mass TRanslator into Pathways

This is MassTRIX reloaded, the 3rd version of MassTRIX.

The jobs on the old server remain still available at this link.
Should you encounter any unexpected behaviour, please let us know!

Welcome to MassTRIX!

MassTRIX annotates metabolites in high precision mass spectrometry data.
Metabolite Identification
LC-MS MetID - *Metabolite / Ion annotation*

<table>
<thead>
<tr>
<th>Database name</th>
<th>Compounds found</th>
<th>Total database entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Abstracts (CAS)</td>
<td>181</td>
<td>24,000,000</td>
</tr>
<tr>
<td>Böhlstein Database (MDL)</td>
<td>166</td>
<td>8,000,000</td>
</tr>
<tr>
<td>Dictionary of Natural Products (DNP)</td>
<td>129</td>
<td>170,000</td>
</tr>
<tr>
<td>PubChem (NIH)</td>
<td>19</td>
<td>800,000</td>
</tr>
<tr>
<td>Available Chemicals Directory (MDL)</td>
<td>6</td>
<td>400,000</td>
</tr>
<tr>
<td>ChemIDplus (NIH)</td>
<td>6</td>
<td>370,000</td>
</tr>
<tr>
<td>KEGG (Kyoto University)</td>
<td>3</td>
<td>13,000</td>
</tr>
<tr>
<td>NIST05 (NIST mass spectral database)</td>
<td>2</td>
<td>163,000</td>
</tr>
<tr>
<td>MOLGEN molecular isomer generator</td>
<td>788,000</td>
<td>-</td>
</tr>
</tbody>
</table>

(allowing 2 benzene groups; 1 ether group, 1 keto group; 5 hydroxy groups)

Kind & Fiehn, BMC Bioinformatics 2006, 7:234
Metabolite Identification
LC-MS MetID - *Metabolite / Ion annotation*

Kind & Fiehn, BMC Bioinformatics 2006, 7:234
Metabolite Identification
LC-MS MetID - Metabolite / Ion annotation

Table 3: Number of possible molecular formulas at different levels of mass accuracy and the impact of isotopic abundance accuracy. A mass spectrometer capable of 3 ppm but with 2% correct isotopic pattern outperforms even a (non-existing) mass spectrometer with 0.1 ppm mass accuracy! The results are computed for randomly selected targets, so single results vary but the trend remains. LEWIS and SENIOR check was applied. Candidates with unrelated high element counts were already excluded.

<table>
<thead>
<tr>
<th>Molecular mass [Da]</th>
<th>without isotope abundance information</th>
<th>2% isotopic abundance accuracy</th>
<th>5% isotopic abundance accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 ppm</td>
<td>5 ppm</td>
<td>3 ppm</td>
</tr>
<tr>
<td>150</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>300</td>
<td>24</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>400</td>
<td>78</td>
<td>37</td>
<td>23</td>
</tr>
<tr>
<td>500</td>
<td>266</td>
<td>115</td>
<td>64</td>
</tr>
<tr>
<td>600</td>
<td>505</td>
<td>257</td>
<td>155</td>
</tr>
<tr>
<td>700</td>
<td>1046</td>
<td>538</td>
<td>321</td>
</tr>
<tr>
<td>800</td>
<td>1964</td>
<td>973</td>
<td>599</td>
</tr>
<tr>
<td>900</td>
<td>3447</td>
<td>1712</td>
<td>1045</td>
</tr>
</tbody>
</table>

Kind & Fiehn, BMC Bioinformatics 2006, 7:234
Metabolite Identification

LC-MS MetID - Metabolite / Ion annotation
Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*

- MS signals from the same metabolite co-elute and co-vary

- Correlation analysis can be used to identify in-source fragments, multimers

- The yielded mass spectra can be used for search in different spectral databases

- Based on the same approach, indiscriminant MS/MS can be used to generate fragmentation for every eluting compound

Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • *in silico* approaches

• *de novo* identification
Metabolite Identification

LC-MS MetID – Formula calculation

Accurate Isotopic Abundance
Accurate Mass

Seven Golden Rules
- Element number restrictions
- Lewis and Senior check
- Isotopic pattern filter
- Hydration/mass ratio check
- NOPs ratio check
- Bement probability check
- TMS check

CSLS service (open access)
30 million compound database

Kind & Fiehn, BMC Bioinformatics 2007, 8:105
Metabolite Identification

LC-MS MetID – Formula calculation

1. Apply heuristic restrictions for number of elements during formula generation
2. Perform LEWIS and SENIOR check
   - LEWIS rule: In its simplest form, the LEWIS rule demands that molecules consisting of main group elements, especially carbon, nitrogen and oxygen, share electrons in a way that all atoms have completely filled s, p-valence shells ('octet rule').
   - SENIOR rule: i) The sum of valences or the total number of atoms having odd valences is even. ii) The sum of valences is greater than or equal to twice the maximum valence. iii) The sum of valences is greater than or equal to twice the number of atoms minus 1.
3. Perform isotopic pattern filter
4. Perform H/C ratio check (hydrogen/carbon ratio)
5. Perform NOPS ratio check (N, O, P, S/C ratios)
6. Perform heuristic HNOPS probability check (H, N, O, P, S/C high probability ratios)
7. Perform –TMS check (for GC-MS if a silylation step is involved)

Kind & Fiehn, BMC Bioinformatics 2007, 8:105
Metabolite Identification

LC-MS MetID – Formula calculation

- Formula calculation can be enhanced by using results from MS/MS
- Sum of fragment sum formula +/- neutral losses = parent sum formula
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation / seven golden rules
    • Utilization of isotopes
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • *in silico* approaches

• *de novo* identification
Metabolite Identification

**LC-MS MetID – Utilization of isotopes**

Stupp et al., Anal. Chem. 2013, 85, 11858–11865
Metabolite Identification

LC-MS MetID – Utilization of isotopes

Stupp et al., Anal. Chem. 2013, 85, 11858−11865
Metabolite Identification

**LC-MS MetID – Utilization of isotopes**

Patrick Giavalisco et al., The Plant Journal (2011) 68, 364–376
Metabolite Identification

LC-MS MetID – Utilization of isotopes

4-Methylthiobutyl glucosinolate (Glucoerucin) 
\([M - H^+]\)

\[ \text{C}_{12} \text{H}_{23} \text{N}_1 \text{O}_9 \text{S}_3 \]

Average RT shift 0.6 sec

Patrick Giavalisco et al., The Plant Journal (2011) 68, 364–376
Metabolite Identification

LC-MS MetID – Utilization of isotopes

Patrick Giavalisco et al., The Plant Journal (2011) 68, 364–376
Metabolite Identification

LC-MS MetID – Utilization of isotopes

T. Nagao et al., Analytica Chimica Acta 813 (2014) 70–76
Metabolite Identification

LC-MS MetID – Utilization of isotopes

T. Nagao et al., Analytica Chimica Acta 813 (2014) 70–76
## Metabolite Identification

**LC-MS MetID – Utilization of isotopes**

<table>
<thead>
<tr>
<th>Element</th>
<th>Isotope</th>
<th>Abundance (%)</th>
<th>Mass</th>
<th>Massdifference</th>
</tr>
</thead>
<tbody>
<tr>
<td>H, hydrogen</td>
<td>$^1\text{H}$</td>
<td>99.985</td>
<td>1.007825</td>
<td></td>
</tr>
<tr>
<td>C, carbon</td>
<td>$^{12}\text{C}$</td>
<td>98.93</td>
<td>12.000000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$^{13}\text{C}$</td>
<td>1.07</td>
<td>13.003355</td>
<td></td>
</tr>
<tr>
<td>N, nitrogen</td>
<td>$^{14}\text{N}$</td>
<td>99.632</td>
<td>14.003074</td>
<td>0.997035</td>
</tr>
<tr>
<td></td>
<td>$^{15}\text{N}$</td>
<td>0.368</td>
<td>15.000109</td>
<td></td>
</tr>
<tr>
<td>O, oxygen</td>
<td>$^{16}\text{O}$</td>
<td>99.757</td>
<td>15.994915</td>
<td>1.004217</td>
</tr>
<tr>
<td></td>
<td>$^{17}\text{O}$</td>
<td>0.038</td>
<td>16.999132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$^{18}\text{O}$</td>
<td>0.205</td>
<td>17.999160</td>
<td>2.004245</td>
</tr>
<tr>
<td>P, phosphorus</td>
<td>$^{31}\text{P}$</td>
<td>100</td>
<td>30.973762</td>
<td></td>
</tr>
<tr>
<td>S, sulfur</td>
<td>$^{32}\text{S}$</td>
<td>94.93</td>
<td>31.972071</td>
<td>0.999388</td>
</tr>
<tr>
<td></td>
<td>$^{33}\text{S}$</td>
<td>0.76</td>
<td>32.971459</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$^{34}\text{S}$</td>
<td>4.29</td>
<td>33.967867</td>
<td>1.995796</td>
</tr>
<tr>
<td>Cl, chlorine</td>
<td>$^{35}\text{Cl}$</td>
<td>75.78</td>
<td>34.968853</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$^{37}\text{Cl}$</td>
<td>24.22</td>
<td>36.965903</td>
<td>1.99705</td>
</tr>
</tbody>
</table>
Metabolite Identification

*LC-MS MetID – Utilization of isotopes*

Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • in silico approaches

• de novo identification
Metabolite Identification

*LC-MS MetID – MS/MS and RT*

- Metabolomics Tandem MS databases
  - Metlin
  - MassBank of Japan
  - MassBank of Europe
  - MassBank of North America (Mona)
  - Global Natural Products Social Molecular Networking (GNPS)
  - Human Metabolome Database (HMDB)
  - RIKEN MSn spectral database (Respect)
- Overlap between all spectra (Splash http://splash.fiehnlab.ucdavis.edu/)
- *in silico tandem* solutions
  - MetFrag / MetFusion
  - CSI:FingerID
  - CFM-ID
  - MAGMa
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

https://gnps.ucsd.edu/ProteoSAFe/gnpsspectrumspectrum.jsp?SpectrumID=CCMSLIB00000079354

http://mona.fiehnlab.ucdavis.edu/spectra/browse
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

![Graph A](image1.png)

![Graph B](image2.png)

![Graph C](image3.png)
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

https://metlin.scripps.edu/index.php

https://metlin.scripps.edu/fragment_search_multi.php
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

![Graph and Table]

**Total: 38 Metabolites**

**Helmholtz Zentrum München**

Deutsches Forschungszentrum für Gesundheit und Umwelt

![Logo]
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

3.69 minutes

4.01 minutes
Metabolite Identification

LC-MS MetID – MS/MS and RT

http://msbi.ipb-halle.de/MetFusion/

Metabolite Identification

**LC-MS MetID – MS/MS and RT**

Metabolite Identification

LC-MS MetID – MS/MS and RT

3.69 minutes

4.01 minutes

Metabolite Identification

**LC-MS MetID – MS/MS and RT**

3.69 minutes

Apigenin

Kaempferol

Luteolin

4.01 minutes

Kaempferol

Maritenein

Fisetin

Luteolin

Datiscetin

Metabolite Identification

LC-MS MetID – MS/MS and RT

Metabolite Identification

LC-MS MetID – MS/MS and RT
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

- RTs can be projected between different (comparable) LC-MS setups
- If the same compounds have been measured on both systems, a function for projection can be calculated allowing to predict RT of compound measured on one system to another

Stanstrup et al., Anal. Chem., 2015, 87 (18), pp 9421–9428
Metabolite Identification

**LC-MS MetID – MS/MS and RT**

http://predret.org/how-to-use-predret/
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • in silico approaches

• de novo identification
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
    • in silico approaches
  • de novo identification
New approaches for MetID

Ion mobility
New approaches for MetID

Ion mobility

- IM experiments take usually a few milliseconds
- This makes IM ideal for combination with chromatography and MS
- UPLC peaks have a typical duration of a few seconds, IM needs a few ms and the MS scans in the \( \mu s \) range
- \[ K_0 = \frac{(18\pi)^{1/2}}{16} \frac{z}{(k_b T)^{1/2}} \left[ \frac{1}{m_I} + \frac{1}{m_B} \right]^{1/2} \frac{1}{N \Omega} \]
- Example: Lipid analysis on a Agilent 6560 IMS-QToF
New approaches for MetID

Ion mobility
Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification
• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT
• New approaches for MetID
  • Ion mobility
    • in silico approaches
• de novo identification
New approaches for MetID

_in silico approaches_

- *in silico* approaches do not rely on previously acquired MS/MS for the exact molecule, but either matches predicted fragments from simple bond breaking or rule based fragmentation

- A selection of *in silico* tandem solutions are
  - MetFrag / MetFusion (IPB Halle)
  - CSI:FingerID (University of Jena)
  - CFM-ID (Wishart Lab)
  - MAGMa (Netherlands Metabolomics Center)
New approaches for MetID

*in silico* approaches - *MetFrag*

### Peaks

<table>
<thead>
<tr>
<th>m/z</th>
<th>intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>119.051</td>
<td>45</td>
</tr>
<tr>
<td>123.044</td>
<td>36</td>
</tr>
<tr>
<td>147.044</td>
<td>606</td>
</tr>
<tr>
<td>153.019</td>
<td>999</td>
</tr>
<tr>
<td>179.036</td>
<td>13</td>
</tr>
<tr>
<td>189.058</td>
<td>16</td>
</tr>
<tr>
<td>273.076</td>
<td>999</td>
</tr>
</tbody>
</table>

### MetFrag

1. Query
2. Fragmentation
3. Scoring

- m/z and intensity
- Bond dissociation energy
- Neutral loss

### Exact Mass

272.06847
(or molecular formula)

### Compound Database

- KEGG
- PubChem
- Chemspider

### Ranked Candidates
New approaches for MetID

\textit{in silico} approaches - MetFrag

New approaches for MetID

*in silico* approaches - MetFrag

Outline

A path to metabolite identification

• Introduction
  • Analytical Technologies
  • Terminology in metabolite identification

• Metabolite identification
  • GC-MS MetID
    • EI spectral deconvolution
    • GC-APCI and exact mass
  • LC-MS MetID
    • Metabolite / Ion annotation
    • Formula calculation
    • Utilization of isotopes for formula calculation
    • MS/MS and RT

• New approaches for MetID
  • Ion mobility
  • in silico approaches

• de novo identification
Metabolite Identification

de novo identification of unknowns

• Complete de novo identification is a tedious process and involves several steps

• de novo identification should be only conducted if metabolite cannot be identified by any other means

• Complete de novo identification requires
  • Full structural elucidation
  • Synthesis of a pure reference substance
**de novo identification**

**Fractionation strategies**

- Enough biomass available?
- Potential activity of metabolite, e.g. hormone?
- Test system available?
- Simple or activity guided fractionation
de novo identification

Fractionation strategies

- Mahanti et al. searched for novel ligands of the nuclear hormone receptor DAF-12 in *C. elegans*
- Activity guided fractionation was used in combination with 2D-NMR based metabolomics to find bioactive molecules
- Activity was checked *in vivo* and *in vitro*
- Previously known molecules Δ4- and Δ7-Dafachronic acid were re-identified together with novel variants
- Synthesis of chemical reference standards confirmed identity
de novo identification

Fractionation strategies

Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014
**de novo** identification

*Fractionation strategies*

Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014
de novo identification
Fractionation strategies

Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014
de novo identification

Fractionation strategies
de novo identification
Fractionation strategies

Sumner et al. & Bruker
**de novo** identification

*Fractionation strategies*

---

*Prospect 2 Solid Phase Extraction Unit with 2 x 96 cartridges & deuterated solvent elution*

*Waters Acquity IClass UHPLC with photodiode array detector*

---

*Bruker impact Q TOF*

*Dr. Dennis Fine*

---

*Gilson liquid handler for SPE eluent collection*

---

*Sumner et al. & Bruker*
de novo identification
in silico strategies
de novo identification

in silico strategies
**de novo** identification

**Reporting**

- Please report your de novo identified metabolite in (electronically) searchable manner

Thank you! Questions?