Metabolomics Workflow: From Data Acquisition to Data Analysis

David A Weil, Ph.D.
LC/MS Applications Scientist
Agilent Technologies,
Schaumburg, IL
Agilent Solutions For Metabolomics
Metabolomics is…

Cytoplasm

(Amino acids)

(transamination to
α-ketoglutarate)

Aspartate

Glutamine (from
extrahepatic tissues)

Glutamine

Glutamate dehydrogenase

Mitochondrial
matrix

Urea cycle

Carbamoyl phosphate synthetase I

Argininosuccinate synthetase

Argininosuccinate lyase

Arginase

HNO

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC

COO

NH3

H2N

OOC
A change in one metabolite causes changes in N number of other metabolites. Since biological pathways are interconnected, state to state changes typically result in 10’s to 100’s metabolites changing and not a single metabolite. The quenching of metabolic is critical to measure these subtle differences in metabolite levels.

Metabolomics studies also require suitable sets of biological replicates in order to find the statistically significant changes.

Must use Statistical Analysis to uncover changes that are Random (from biological diversity) versus those highly significant changes.
Targeted versus Global Metabolomics

Targeted - Known metabolites only

- Absolute quantitation/Validation—Need internal standards
- Unit mass instruments (LC-QQQ or GC-QQQ)
- Data is acquired in SIM or MRM mode
- Flux analysis
- Pathway analysis and model building software for biological interpretation

Global Data Acquisition – Unknown Metabolites

- Accurate Mass instrument (LC-TOF, Q-TOF) High Res accurate mass
- Track metabolites using retention time and mass
- Find differential metabolites (features) using statistical analysis programs
- Identify differential metabolites
- Interpret results in context of biological pathway analysis software tools
- Integrate into Systems Biology context
Data Analysis Approaches

Untargeted Data Analysis

- Profiling of all metabolites: Discovery based approach
- Naïve data mining with few assumptions
- Track metabolites using mass and retention time
- Finds unexpected compounds
- Metabolites need to be identified and confirmed after analysis
- Pathway analysis and Biological interpretation

Pathway Targeted Data Analysis

- Mine only pathway metabolites: Hypothesis based approach
- Targeted data mining uses metabolite database
- Track metabolites using metabolite name
- More selective and sensitivity than untargeted analysis
- Metabolites need to be confirmed after analysis
- Pathway analysis and Biological interpretation

One analysis does not preclude doing the other
### Variety of Applications for Statistical Analysis

- Metabolomics
- Proteomics
- Food safety
- Environmental
- Forensics
- Toxicology
- Petrochemical
- Biofuels
Challenges of Metabolomics ➔ Total Workflow

All Steps Critical from Sample Preparation to Identification
Agilent Metabolomics Workflow

Separate & Detect
- GC/MSD
- GC-QQQ
- GC/QTOF

Feature Finding Quantitate
- MassHunter Qual
  - AMDIS or Find by chromatographic deconvolution
- MassHunter Qual
  - MFE, Find by Formula, Find by Ion

Alignment & Statistics

Identify

Pathways

GCMS

LCMS

GC-MSD
GC-QQQ
GC/QTOF

LC-TOF/QTOF
LC-QQQ

Mass Profiler Professional (MPP)
ID Browser
Pathway Architect
81% of respondents say challenges due to **software** NOT **hardware**

1. Compound Identification – 35%
2. Biological Significance – 27%
3. Data Processing – 14%
4. Statistical Analysis – 5%

2009 ASMS Metabolomics Survey Results
http://metabolomics.us/2009/ASMSMetabolomicsWorkshop/SurveyResults/
Instrumentation for Metabolomics

- 7000B GC/QQQ
- 5975C GC/MS
- Hi-DEF Q-TOF 6500 series
- 7700 GC/QTOF
- Infinity 1290 LC
- QQQ 6400 Series
- TOF 6200 series
Sample Preparation: Critical Step First Step

Discovery Metabolomics “Snapshot” of Metabolic Activity and Comparison Between Identical States at Different Conditions

- Metabolic Process on sub-ms time scales
  - Must Rapidly Quench Enzymatic Activity in Cell/Tissue

- Recovery across broad range of metabolites!
  - No Standards-Internal Standard QC
  - Extract Protocols: pH, solvent composition, Temperature
  - Stability: Remove Proteins, Chemical Degradation (-80°C)
  - Fractionation reduce ion suppression

Objective: Develop a Universal Chromatographic Method that Limits Co-Elution, Ion Suppression, Peak Tailing, Over Diverse Compound Classes...

Reality: No One Method!!

Volutility: GC vs LC vs CE

Polarity: Acids, Bases, NonPolar, Aromatic

Hydrophobic vs Hydrophilic

Matrix Effects: Ion Suppression Concentration Ranges

LC Method: Reverse Phase, Ion Pair, HILIC, RP, Mixed Mode, ANP

Challenges for Detection of Anionic Metabolites

- **Sticking to Metal Surfaces:**

- **Time Consuming** LC cleaning Procedure: nitric acid, phosphoric acid, IPA, Water, flushing solution then still need to flush with 6 mM of EDTA to reduce the binding of phosphorylate metabolites to all the metal sites.

- Mass Spectrometer: **Metal Needles** 308 Stainless Steel (Copper Isotopes) replace with new 316 SS Needles but they contain Fe (unique isotopes). Must also rinse the nebulizer with EDTA!!!

- Replace Stainless Steel Capillary tubing with PEEK Tubing but it’s not compatible at high pressures.

- Stainless Steel Clad PEEK Capillaries (in Bioinert)

- Lots of References about this!
  
  
1260 Infinity Bio-inert HPLC

What is similar compared to standard 1260 Infinity Quaternary LC?

- Pumping performance, gradient performance
- Pumping Specifications
- Detector, Autosampler Performance and Specs
- Overall System Robustness
- TCC specs

600 bar also for fast or high resolution Bioseparations

Modular flexibility maintained
New Capillary Design Ensures Bio-inertness

Capillaries:

Metal cladded PEEK capillary

New capillary technology enables 600 bar
AND is completely metal free !!

Needle:

SST housing for robustness

PEEK and ceramic for metal free sample path
UHPLC Chromatography
Narrower Peaks $\Rightarrow$ Sensitivity, Resolution, Faster

Urine Sample 1.0 ml/min
Water/Acetonitrile Grad

> 450 Features Found
Coupled with Fast Acquisition MS to 40 spectra/s

Peak Width 2s

3-methylxanthine
-0.11 ppm

benzocaine

Flow Rate 1mL/min
MS 10 spectra/s
Ion-Pair Reverse Phase Chromatography

- **Disadvantages:**
  - Primary for Anionic Polar Metabolites
  - Potential Ion Suppression
  - High Chemical Background from Triakyl Amines (TBA, Tributyl)Amine
  - Different Chromatography for Cationic and Anionic Metabolites RT!

- **Advantages:**
  - Established methodology Central Metabolism (TCA, Glycolysis, pentose phosphate pathways)
  - Targeted not Global Metabolomics
  - Applicable to Nanoscale Chromatography

Separation of Anionic Polar Metabolites
Imtakt Mixture Mixed Mode SW-C18 Bio-inert LC

A: 0.2% Acetic Acid
B: 10/10/80 Water/MeOH/IPA 20mM NH₄AC pH7
Initial 95/5 for 10 min, Ramp to 90%B at 25 min
Flow Rate 350 uL/min
Column: Mixed Mode SW
## Polar Mixture

<table>
<thead>
<tr>
<th>RT</th>
<th>Name</th>
<th>Formula</th>
<th>Mass</th>
<th>Score</th>
<th>Diff (DB, ppm)</th>
<th>Height</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.992</td>
<td>Sorbitol</td>
<td>C6 H14 O6</td>
<td>182.0735</td>
<td>99.47</td>
<td>-1.7</td>
<td>536126</td>
<td>1574361</td>
</tr>
<tr>
<td>2.009</td>
<td>sucrose</td>
<td>C12 H22 O11</td>
<td>342.11673</td>
<td>98.89</td>
<td>-1.52</td>
<td>162139</td>
<td>497005</td>
</tr>
<tr>
<td>2.042</td>
<td>D-Ribose</td>
<td>C5 H10 O5</td>
<td>150.0525</td>
<td>86.82</td>
<td>1.98</td>
<td>12011</td>
<td>38316</td>
</tr>
<tr>
<td>2.075</td>
<td>Hydroxy-L-Proline</td>
<td>C5 H9 N O3</td>
<td>131.05828</td>
<td>99.96</td>
<td>-0.28</td>
<td>467696</td>
<td>1299446</td>
</tr>
<tr>
<td>2.125</td>
<td>cystine</td>
<td>C6 H12 N2 O4 S2</td>
<td>240.02407</td>
<td>98.08</td>
<td>-0.91</td>
<td>167170</td>
<td>556195</td>
</tr>
<tr>
<td>2.159</td>
<td>Alanine</td>
<td>C3 H7 N O2</td>
<td>89.04753</td>
<td>99.23</td>
<td>1.73</td>
<td>57262</td>
<td>167257</td>
</tr>
<tr>
<td>2.159</td>
<td>Cytidine monophosphate (CMP)</td>
<td>C9 H14 N3 O8 P</td>
<td>323.05153</td>
<td>99.48</td>
<td>0.98</td>
<td>117165</td>
<td>513188</td>
</tr>
<tr>
<td>2.175</td>
<td>serine</td>
<td>C3 H7 N O3</td>
<td>105.04242</td>
<td>99.78</td>
<td>1.58</td>
<td>69445</td>
<td>206988</td>
</tr>
<tr>
<td>2.175</td>
<td>Glutamine</td>
<td>C5 H10 N2 O3</td>
<td>146.06881</td>
<td>99.28</td>
<td>2.27</td>
<td>272849</td>
<td>800756</td>
</tr>
<tr>
<td>2.209</td>
<td>Proline</td>
<td>C5 H9 N O2</td>
<td>115.06314</td>
<td>99.71</td>
<td>1.67</td>
<td>179333</td>
<td>509674</td>
</tr>
<tr>
<td>2.225</td>
<td>Glycine</td>
<td>C2 H5 N O2</td>
<td>75.03193</td>
<td>79.36</td>
<td>1.29</td>
<td>2806</td>
<td>9763</td>
</tr>
<tr>
<td>2.225</td>
<td>Threonine</td>
<td>C4 H9 N O3</td>
<td>119.05807</td>
<td>99.63</td>
<td>1.42</td>
<td>136543</td>
<td>364442</td>
</tr>
<tr>
<td>2.275</td>
<td>L-Citrulline</td>
<td>C6 H13 N3 O3</td>
<td>175.09559</td>
<td>99.88</td>
<td>0.6</td>
<td>224326</td>
<td>703696</td>
</tr>
<tr>
<td>2.342</td>
<td>Sarcosine</td>
<td>C3 H7 N O2</td>
<td>89.04773</td>
<td>99.93</td>
<td>-0.55</td>
<td>67172</td>
<td>358428</td>
</tr>
<tr>
<td>2.342</td>
<td>Glutamic Acid</td>
<td>C5 H9 N O4</td>
<td>147.05325</td>
<td>99.95</td>
<td>-0.6</td>
<td>165652</td>
<td>548540</td>
</tr>
<tr>
<td>2.375</td>
<td>Aspartic acid</td>
<td>C4 H7 N O4</td>
<td>133.03742</td>
<td>99.45</td>
<td>0.68</td>
<td>114172</td>
<td>459827</td>
</tr>
<tr>
<td>2.525</td>
<td>Gammaaminobutyric acid</td>
<td>C4 H9 N O2</td>
<td>103.06372</td>
<td>98.51</td>
<td>-3.76</td>
<td>227600</td>
<td>713745</td>
</tr>
<tr>
<td>2.525</td>
<td>Uracil</td>
<td>C4 H4 N2 O2</td>
<td>112.02762</td>
<td>87.13</td>
<td>-3</td>
<td>26614</td>
<td>100316</td>
</tr>
<tr>
<td>2.525</td>
<td>Uridine</td>
<td>C9 H12 N2 O6</td>
<td>244.07039</td>
<td>96.98</td>
<td>-3.47</td>
<td>395024</td>
<td>1357680</td>
</tr>
<tr>
<td>3.025</td>
<td>Valine</td>
<td>C5 H11 N O2</td>
<td>117.07903</td>
<td>99.92</td>
<td>-0.46</td>
<td>264902</td>
<td>891351</td>
</tr>
<tr>
<td>3.142</td>
<td>AMP</td>
<td>C10 H14 N5 O7 P</td>
<td>347.06317</td>
<td>99.24</td>
<td>-0.25</td>
<td>67044</td>
<td>409471</td>
</tr>
<tr>
<td>3.192</td>
<td>methionine</td>
<td>C5 H11 N O2 S</td>
<td>149.05092</td>
<td>99.53</td>
<td>0.85</td>
<td>170962</td>
<td>522112</td>
</tr>
<tr>
<td>3.192</td>
<td>Guanine</td>
<td>C5 H5 N O5</td>
<td>151.04811</td>
<td>87.66</td>
<td>8.61</td>
<td>8519</td>
<td>34445</td>
</tr>
<tr>
<td>3.209</td>
<td>Succinic acid</td>
<td>C4 H6 O4</td>
<td>118.02642</td>
<td>99.74</td>
<td>1.64</td>
<td>354931</td>
<td>1731122</td>
</tr>
<tr>
<td>3.225</td>
<td>Guanosine</td>
<td>C10 H13 N5 O5</td>
<td>283.09136</td>
<td>99.12</td>
<td>1.02</td>
<td>88723</td>
<td>394126</td>
</tr>
<tr>
<td>RT</td>
<td>Name</td>
<td>Formula</td>
<td>Mass</td>
<td>Score</td>
<td>Diff (DB, ppm)</td>
<td>Height</td>
<td>Area</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>-------</td>
<td>----------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>3.225</td>
<td>Guanosine</td>
<td>C10 H13 N5 O5</td>
<td>283.09138</td>
<td>99.12</td>
<td>0.10</td>
<td>88723</td>
<td>394126</td>
</tr>
<tr>
<td>4.375</td>
<td>Guanosine 5-monophosphate</td>
<td>C10 H14 N5 O8 P</td>
<td>363.05763</td>
<td>99.86</td>
<td>0.47</td>
<td>26068</td>
<td>314974</td>
</tr>
<tr>
<td>4.425</td>
<td>Tyrosine</td>
<td>C9 H11 N O3</td>
<td>181.07402</td>
<td>99.82</td>
<td>-0.7</td>
<td>97119</td>
<td>360482</td>
</tr>
<tr>
<td>4.425</td>
<td>Uridine monophosphate (UMP)</td>
<td>C9 H13 N2 O9 P</td>
<td>324.03611</td>
<td>99.75</td>
<td>-0.73</td>
<td>101572</td>
<td>705287</td>
</tr>
<tr>
<td>4.625</td>
<td>Pyruvic acid</td>
<td>C3 H4 O3</td>
<td>88.01599</td>
<td>99.7</td>
<td>0.53</td>
<td>28037</td>
<td>482620</td>
</tr>
<tr>
<td>4.975</td>
<td>Leucine</td>
<td>C6 H13 N O2</td>
<td>131.0942</td>
<td>98.82</td>
<td>3.29</td>
<td>213238</td>
<td>1264367</td>
</tr>
<tr>
<td>5.075</td>
<td>1-methyl-L-histidine</td>
<td>C7 H11 N3 O2</td>
<td>169.08492</td>
<td>99.67</td>
<td>1.27</td>
<td>168731</td>
<td>1142587</td>
</tr>
<tr>
<td>5.342</td>
<td>Cytosine</td>
<td>C4 H5 N O3</td>
<td>111.04331</td>
<td>87.62</td>
<td>-0.41</td>
<td>4296</td>
<td>24044</td>
</tr>
<tr>
<td>5.425</td>
<td>Leucine</td>
<td>C6 H13 N O2</td>
<td>131.09485</td>
<td>99.68</td>
<td>-1.67</td>
<td>221774</td>
<td>1414527</td>
</tr>
<tr>
<td>5.475</td>
<td>Histidine</td>
<td>C6 H9 N3 O2</td>
<td>155.0697</td>
<td>99.61</td>
<td>-1.45</td>
<td>135529</td>
<td>960946</td>
</tr>
<tr>
<td>5.542</td>
<td>Maleic acid</td>
<td>C4 H4 O4</td>
<td>116.0114</td>
<td>97.78</td>
<td>-3.75</td>
<td>52387</td>
<td>348076</td>
</tr>
<tr>
<td>5.992</td>
<td>Adenine</td>
<td>C5 H5 N5</td>
<td>135.05444</td>
<td>99.82</td>
<td>0.43</td>
<td>61060</td>
<td>463873</td>
</tr>
<tr>
<td>5.992</td>
<td>Adenosine</td>
<td>C10 H13 N5 O4</td>
<td>267.0962</td>
<td>99.85</td>
<td>-0.65</td>
<td>31149</td>
<td>226696</td>
</tr>
<tr>
<td>6.675</td>
<td>Lysine</td>
<td>C6 H14 N2 O2</td>
<td>146.10567</td>
<td>99.79</td>
<td>-0.95</td>
<td>112758</td>
<td>986735</td>
</tr>
<tr>
<td>6.792</td>
<td>Phenylalanine</td>
<td>C9 H11 N O2</td>
<td>165.07907</td>
<td>99.9</td>
<td>-0.57</td>
<td>187739</td>
<td>1590659</td>
</tr>
<tr>
<td>8.458</td>
<td>Fumaric acid</td>
<td>C4 H4 O4</td>
<td>116.0109</td>
<td>99.88</td>
<td>0.49</td>
<td>104032</td>
<td>1533610</td>
</tr>
<tr>
<td>8.642</td>
<td>Arginine</td>
<td>C6 H14 N4 O2</td>
<td>174.11162</td>
<td>99.84</td>
<td>0.36</td>
<td>78736</td>
<td>1036645</td>
</tr>
<tr>
<td>11.408</td>
<td>Aconitic acid</td>
<td>C6 H6 O6</td>
<td>174.01623</td>
<td>86.08</td>
<td>1.19</td>
<td>16155</td>
<td>289391</td>
</tr>
<tr>
<td>16.092</td>
<td>Sulphamethazine</td>
<td>C12 H14 N4 O2 S</td>
<td>278.0845</td>
<td>97.8</td>
<td>-2.69</td>
<td>27780</td>
<td>213906</td>
</tr>
<tr>
<td>16.675</td>
<td>Tryptophan</td>
<td>C11 H12 N2 O2</td>
<td>204.0905</td>
<td>98.09</td>
<td>-3.04</td>
<td>138678</td>
<td>611801</td>
</tr>
<tr>
<td>17.142</td>
<td>Sulphamethizole</td>
<td>C9 H10 N4 O2 S 2</td>
<td>270.02533</td>
<td>97.33</td>
<td>-2.98</td>
<td>149321</td>
<td>758143</td>
</tr>
<tr>
<td>18.125</td>
<td>Sulphachloropyridazine</td>
<td>C10 H9 Cl N4 O2 S</td>
<td>284.01421</td>
<td>97.77</td>
<td>-2.59</td>
<td>288612</td>
<td>1329831</td>
</tr>
<tr>
<td>19.658</td>
<td>Sulphadimethoxime</td>
<td>C12 H14 N4 O4 S</td>
<td>310.07373</td>
<td>99.65</td>
<td>-0.5</td>
<td>452327</td>
<td>1959434</td>
</tr>
<tr>
<td>19.991</td>
<td>Sulphadimethoxime</td>
<td>C12 H14 N4 O4 S</td>
<td>310.07411</td>
<td>98.88</td>
<td>-1.71</td>
<td>445793</td>
<td>2021281</td>
</tr>
<tr>
<td>24.591</td>
<td>Adenine</td>
<td>C5 H5 N5</td>
<td>135.05417</td>
<td>83.62</td>
<td>2.41</td>
<td>7385</td>
<td>118219</td>
</tr>
</tbody>
</table>
ANP Chromatography: Effect Of Trace Sodium Leaching From Glass Bottles

Glass Bottles
Fresh Solvent

Glass Bottles
After Standing for a Few Days
Aqueous Normal Phase Separation Method (ANP)
UDP-Glucose And UDP-Galactose (C_{15}H_{24}N_{2}O_{17}P_{2})

**Column:** MicroSolv Technology Corp
Cogent Diamond Hydride 2.1 x 150 mm

**Solvents:**
Bottle A: **Strong Solvent**
50% MeOH/ 50% H_{2}O / 0.05% FA
Bottle B: **Weak Solvent**
90% ACN/10% H_{2}O with 10 mM NH_{4} Acetate

Flow Rate: 0.4 ml/min
Stop time: 20 min
Post time: 5 min
Tandem LCMS ANP with RP
Metabolomics group in Australia
Tandem LCMS ANP with RP
QTOF Innovations from Agilent

- Ion Beam Compression Technology (resolution + mass accuracy)
- Orthogonal spray source (signal-to-noise)
- iFunnel technology
- Dual-stage ion mirror (resolution)
- Longer flight tube (resolution)
- INVAR flight tube (mass accuracy)
- ADC (dynamic range), w/ dual amplifier
- 4 GHz electronics (resolution, mass accuracy, sensitivity, dynamic range)
- 50 Spectra/sec
- Ion acceleration in hexapole collision cell (faster MS/MS spectra)
Hexabore Atmospheric Sampling….  

- **SIX** bores  
- **HALF** as long  

*Six bores, Half the restriction means…*

- 6 times the amount of atmospheric gas sampled  

**AND**  

- 10X the number of ions sampled over wide mass range.  

*But how do we handle all the extra gas molecules?*
Two Stage Ion Funnel Manages High Gas Load

Stage 1 offset deflects the high quantity of gas exiting the hexabore capillary.
How is an Ion Funnel Constructed?

Previously, with many metal plates….

… resulting in a large capacitive load and relatively large RF power electronics.

The 6550 design uses printed circuit board technology with only a small conductive rim resulting in a reduced capacitance load. This makes +/- ion switching possible.
Data Analysis Workflow Methods

Untargeted
- Separate & Detect
  - LC-TOF/QTOF
  - MassHunter Qual
    - Molecular Feature Extraction

Targeted
- Separate & Detect
  - LC-TOF/QTOF
  - MassHunter Qual
    - Find by Formula

- Feature Finding
  - MassHunter Qual
  - Mass Profiler Professional

- Alignment & Statistics
  - Mass Profiler Professional

- Identify
  - ID Browser

- Pathways
  - Pathway Architect
Untargeted Metabolomics Data Analysis Workflow

Run Sample & Control in MS mode → Find Compounds - Molecular Feature Extraction (MFE) → Statistically Filter Compounds - Guided/Advanced Workflow - Recursive Analysis → Identify Compounds (Formula Level) - AMRT Database Search - Molecular Formula Generator (MFG)

MassHunter Acquisition → MassHunter Qualitative Analysis → MassHunter Mass Profiler Professional → MassHunter ID Browser

Run Sample in MS/MS mode → Find Compounds - MFE - Auto or Targeted MS/MS → Identify Compounds - AM Spectral Library Search - MFG (MS/MS) → Identify Compounds (Structure Level) - MFG (MS/MS) - Molecular Structure Correlation (MSC)

MassHunter Acquisition → MassHunter Qualitative Analysis → MassHunter Qualitative Analysis → MassHunter Molecular Structure Correlation
Molecular Feature Extraction (MFE)
Automated Data Reduction Software

Finds Features in TOF/QTOF Data

Data Reduced sum intensities of isotopes, adducts, clusters and multiply charges ions together.
Molecular Feature Extraction (MFE)  
Unsupervised Data Mining Software

Data is processed by proprietary feature finding algorithm

• Find chromatographic peaks
  – Find all ions that are related
  – Include any adducts, such as Na\(^+\) or K\(^+\)
  – Include isotopes ([M+H]\(^+\), [M+H+1]\(^+\),…
  – Check for dimers
  – Create a compound chromatograms (ECC) and spectra

• Sum all ion signals into one value (Feature)

• Fully automated processing
Pathways to PCDL

Convert pathway metabolite information into Agilent personal compound databases

- Pathway database source - WikiPathways, BioCyc and KEGG
- Select one to many pathways
- Removes redundant metabolites
- Adds compound information – Formula, Compound ID(s), Name, Structure

Can link to METLIN PCDL to add compound information
- Retention time or MS/MS spectra
Power of Pathway Database

Mine Data Using Find by Formula and Database

- Extract chromatogram and spectra using empirical formula and user settable rules
- Use a metabolite database created by Pathways to PCDL
- Create Find by Formula method
- Chromatograms extracted and integrated
- Spectra scored from empirical formula
  - Isotope mass values
  - Isotope ratio
- User specifies match criteria threshold
  - Spectra score
  - Retention time (optional) increases specificity
- User can review and edit results
- Produce a CEF file for import into MPP

Targeted Data Mining of Qualitative Data for Greater Specificity
MFG Scoring Algorithm

**MS Level Scoring:** Takes into account mass accuracy (MS), isotopic abundance and isotopic spacing.

- Monoisotopic mass (varies in ppm)
- Isotope spacing (varies in ppm)
- Isotope distribution (varies in %)

\[
\text{MFG Score (MS)} = \frac{\text{Monoisotopic mass}}{\text{Isotope spacing}} + \frac{\text{Isotope distribution}}{100}
\]

**MS/MS Level Scoring:** Takes into account mass accuracy (MS/MS) of fragment ions and neutral loss info.

**MFG Score, Overall** = MFG Score (MS) + MFG Score (MS/MS)
Mass Error < 0.1 ppm for L-Tryptophan C₁₁H₁₂N₂O₂

-0.1 ppm Error
Accurate Mass MS/MS Examples

Glutathione C$_{11}$H$_{14}$N$_{7}$SO$_{2}$
Methyl Adenosine $\text{C}_{11}\text{H}_{15}\text{N}_5\text{O}_4$ -1.68 ppm

Cpd 277: 2.225 281.1129 1-Methyladenosine; C11 H15 N5 O4; +ESI Scan (2.209, 2.226 min, 2 Scans) Frag=330.0 V Sample 10 1uL.d

- 130.15936
- 114.09140
- 141.95851
- 172.97701
- 227.17592

Cpd 277: 2.225 281.1129 1-Methyladenosine; C11 H15 N5 O4; +ESI Product Ion (2.218, 2.232 min, 2 Scans) Frag=330.0 V CID@17.9 (282.12015[z=1] -> **) Sample 10 1uL.d AvgCE

- 103.03786 C2 H5 N3 O2
- 170.08129 C8 H12 N O3
- 150.07755 C6 H8 N5

Counts vs. Mass-to-Charge (m/z)

<table>
<thead>
<tr>
<th>m/z</th>
<th>Formula</th>
<th>Intensity</th>
<th>Error</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>170.08129</td>
<td>C8 H12 N O3</td>
<td>13.4</td>
<td>-0.7</td>
<td>112.0385</td>
</tr>
<tr>
<td>150.07755</td>
<td>C6 H8 N5</td>
<td>69.4</td>
<td>-0.83</td>
<td>132.0423</td>
</tr>
<tr>
<td>103.0737</td>
<td>C3 H9 N3 O2</td>
<td>3.8</td>
<td>3.03</td>
<td>179.0457</td>
</tr>
<tr>
<td>103.03786</td>
<td>C2 H5 N3 O2</td>
<td>7.0</td>
<td>-2.25</td>
<td>179.0821</td>
</tr>
</tbody>
</table>
Increasing Your Confidence in Compound Identification

Compounds can be identified:

1. Database matching using accurate mass measurement
2. Database matching with isotope pattern matching
3. Database matching with isotope pattern matching and retention time
4. MS/MS library matching
5. MS/MS library and retention time matching

Citric acid pathway map
METLIN PCDL **accurate mass** LC-MS/MS

- MS database contains > 64092 compounds
- MS/MS spectra from mono-isotopic ion, collected in ESI pos. and neg. modes and at three collision energies: 10, 20 and 40eV
- MS/MS spectra are curated for quality
  - Fragment ions are confirmed
  - Fragment ions are mass corrected
  - Noise ions removed
  - Manually reviewed
- MS/MS Library contains > 8040 compounds

**Also allows customers to build their own databases and libraries**
IDBrowser Identification of the 6 Compounds

<table>
<thead>
<tr>
<th>Label</th>
<th>Name</th>
<th>Formula</th>
<th>Notes</th>
<th>CAS</th>
<th>Overall Score</th>
<th>RT</th>
<th>Mass</th>
<th>Mass (DB)</th>
<th>Diff (ppm)</th>
<th>Mass (MFG)</th>
<th>Diff (MFG, ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cpd 2</td>
<td>Isoamyl nitrite</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>89.41</td>
<td>7.08</td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>Isoamyl nitrite</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>89.41</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>5-Aminopentanoic acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>Valine</td>
<td>C5 H11 N O2</td>
<td>Positive MS/MS</td>
<td>72-19-4</td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>N-Methyl-a-aminoisobutyric acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>4-Aminopentanoic acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>2S-Aminopentanoic acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>4R-Aminopentanoic acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>4S-Aminopentanoic acid</td>
<td>C5 H11 N O2</td>
<td></td>
<td></td>
<td>87.98</td>
<td></td>
<td>117.0791</td>
<td>117.079</td>
<td>-1.03</td>
<td>117.079</td>
<td>-1.94</td>
</tr>
<tr>
<td>Cpd 1</td>
<td>Deoxyadenosine</td>
<td>C10 H13 N O3</td>
<td>Endogenous Meta</td>
<td>958-99-9</td>
<td>96.19</td>
<td>3.175</td>
<td>251.1016</td>
<td>251.1016</td>
<td>1.02</td>
<td>251.1018</td>
<td>1.02</td>
</tr>
<tr>
<td>Cpd 5</td>
<td>Cytochrome</td>
<td>C4 H5 N O3</td>
<td>Group A 07/27/04</td>
<td>71-30-7</td>
<td>92.7</td>
<td>4.783</td>
<td>111.0435</td>
<td>111.0433</td>
<td>-2.27</td>
<td>111.0433</td>
<td>-2.27</td>
</tr>
<tr>
<td>Cpd 4</td>
<td>Adenine</td>
<td>C5 H5 N O5</td>
<td>Endogenous Meta</td>
<td>72-24-5</td>
<td>83.59</td>
<td>3.177</td>
<td>135.0545</td>
<td>135.0544</td>
<td>-3.84</td>
<td>135.0545</td>
<td>-3.84</td>
</tr>
<tr>
<td>Cpd 3</td>
<td>C7 H17 N O2</td>
<td></td>
<td></td>
<td></td>
<td>92.96</td>
<td>7.099</td>
<td>175.1316</td>
<td>175.1321</td>
<td>2.49</td>
<td>175.1321</td>
<td>2.49</td>
</tr>
<tr>
<td>Cpd 5</td>
<td>C8 H12 O P2</td>
<td></td>
<td></td>
<td></td>
<td>78.95</td>
<td>26.872</td>
<td>249.9804</td>
<td>249.9805</td>
<td>0.48</td>
<td>249.9805</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Agilent Technologies
Molecular Structure Correlation Software $\text{C}_{14}\text{H}_{20}\text{O}_{4}$
MassHunter Profinder: Batch Extraction and Review

- Standalone program based on MassHunter Qualitative software
- Performs untargeted data mining
  - Uses MFE
  - Automatic recursive analysis
- Performs targeted data mining
  - Uses Find by Formula
- Supports manual review and editing
- Displays four windows of information
  - Extraction results summary
  - Compound details
  - Compound chromatograms
  - Compound spectra

Easy Data Review and Higher Quality Data for MPP
MassHunter Profinder: Batch Extraction and Review

Extraction Summary

<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
<th>Found</th>
<th>Masked</th>
<th>Mass (Tgl)</th>
<th>Mass (med)</th>
<th>RT (med)</th>
<th>RT (open)</th>
<th>RT (max width)</th>
<th>Area (med)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Pyrrole-2-car</td>
<td>C5H7N2O2</td>
<td>4</td>
<td>5</td>
<td>113.0477</td>
<td>113.0477</td>
<td>7.296</td>
<td>0.013</td>
<td>1.879</td>
<td>259564.4</td>
</tr>
<tr>
<td>2-Pyrrole-2-car</td>
<td>C5H7N2O2</td>
<td>5</td>
<td>4</td>
<td>113.0477</td>
<td>113.0479</td>
<td>7.787</td>
<td>0.02</td>
<td>2.332</td>
<td>439540.7</td>
</tr>
<tr>
<td>5-Amino-2-car</td>
<td>C5H9N3O3</td>
<td>3</td>
<td>6</td>
<td>131.0562</td>
<td>131.0561</td>
<td>1.111</td>
<td>0.001</td>
<td>0.556</td>
<td>21347.6</td>
</tr>
<tr>
<td>5-Amino-2-car</td>
<td>C5H9N3O3</td>
<td>0</td>
<td>9</td>
<td>131.0582</td>
<td>131.0583</td>
<td>1.11</td>
<td>0.003</td>
<td>0.543</td>
<td>12983.3</td>
</tr>
</tbody>
</table>

Compound Chromatograms

Compound Details

Designed for Easy Data Review and Visualization
MassHunter Profinder: 
Manual Editing of Data

Designed for Easy Manual Editing of Data Results

Tables Automatically Update

Originally Selected Compounds

Manually Selected Compounds

Designed for Easy Manual Editing of Data Results
Mass Profiler Professional: Statistical Analysis and Visualization Software

- Designed for Mass Spectrometry data from multiple platforms
- Can Import, store, and visualize
  - Agilent LC/MS Q(TOF), and QQQ
  - Agilent GC/MS Quad, QQQ, and QTOF
  - Agilent ICP/MS
  - Generic file format import
- Extensive statistical analyses tools
  - ANOVA, Clustering, PCA, Fold-change, Volcano plots
- ID Browser for compound identification
- Integrated Biology
- Pathway Architect for biological contextualization
Pathway Directed Experiment Creation

Propose new experiments based on pathway analysis

- Re-examine acquired untargeted metabolomics data based on pathway analysis
- Design new experiments (metabolite, protein or genes) based on pathway results interpretation
BioCyc Pathways

• Produced by SRI under the direction of Peter Karp
• Pathway Architect downloads BioCyc database from Agilent server
• BioCyc content
  • Number of species – 2037
• Curated and computationally derived pathways
  • Tier 1 – literature assembled and manually reviewed
  • Tier 2 – computationally generated with moderate review
  • Tier 3 – computationally generated no review

Additional Pathway Source that has a Large and Deep Species Coverage
Pathway Architect 12.5 in MPP 12.5
BioCyc Pathway Support

BioCyc uses BioPAX 2 and 3 format that does not contain drawing information

- Contains only the compound connections
- Rendering program must provide the drawing logic

Pathway Architect uses a new linear logic for rendering – better visualization

BioCyc: noradrenaline and adrenaline degradation
Agilent-BridgeDB: Enhanced Metabolite Mapping

Metabolites Identifiers – more coverage
- KEGG
- MetaCyc
- PubChem
- LMP
- HMDB
- ChEBI
- CAS

Proteins Identifiers:
- Swiss-Prot
- UniProt
- UniProt/TrEMBL

Genes Identifiers:
- Entrez Gene, GenBank, Ensembl
- EC Number, RefSeq, UniGene, HUGO
- HGNC, EMBL

Resolve the Mapping Problem Between Databases
Multi-Omics Example: Metabolomics → Proteomics

- LC/MS-based Protein Abundance Measurements
- LC/MS or GC/MS-based Metabolite Abundance Measurements
- Joint Pathways experiment: metabolomics / proteomics
Breast Cancer Cell Lines Study: Amino Acid Cycle
Color indicates Fold Changes

<table>
<thead>
<tr>
<th>Entity Name</th>
<th>Compound ...</th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>DB</th>
<th>DB ID</th>
<th>Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Aminobutyrate</td>
<td></td>
<td>-1.676</td>
<td>0.144</td>
<td>0.107</td>
<td>KEGG ID</td>
<td>C02220</td>
<td>2-Aminobutyrate</td>
</tr>
<tr>
<td>3-aminovalerate</td>
<td></td>
<td>0.568</td>
<td>-2.439</td>
<td>0.082</td>
<td>KEGG ID</td>
<td>C02229</td>
<td>3-aminovalerate</td>
</tr>
<tr>
<td>S-methylpropane</td>
<td></td>
<td>1.113</td>
<td>-0.214</td>
<td>-0.106</td>
<td>KEGG ID</td>
<td>C00170</td>
<td>S-Methylpropane</td>
</tr>
<tr>
<td>5-hydroxytryptamine</td>
<td></td>
<td>-15.37</td>
<td>0.095</td>
<td>0.171</td>
<td>CAS Number</td>
<td>4350-09-8</td>
<td>5-Hydroxytryptamine</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine 3',5'</td>
<td>0</td>
<td>4.969</td>
<td>16.33</td>
<td>KEGG ID</td>
<td>C00020</td>
<td>Adenosine 3',5'</td>
</tr>
<tr>
<td>Arginine</td>
<td></td>
<td>-0.873</td>
<td>-0.007</td>
<td>0.281</td>
<td>KEGG ID</td>
<td>C00062</td>
<td>Arginine</td>
</tr>
<tr>
<td>L-Aspartate</td>
<td>Aspartic Acid</td>
<td>0.268</td>
<td>-0.23</td>
<td>-0.049</td>
<td>KEGG ID</td>
<td>C00049</td>
<td>Aspartic Acid</td>
</tr>
<tr>
<td>Creatine</td>
<td></td>
<td>0.137</td>
<td>-0.006</td>
<td>-0.059</td>
<td>KEGG ID</td>
<td>C00300</td>
<td>Creatine</td>
</tr>
<tr>
<td>Creatinine</td>
<td>Creatinine</td>
<td>0.667</td>
<td>-0.246</td>
<td>-0.311</td>
<td>KEGG ID</td>
<td>C00791</td>
<td>Creatinine</td>
</tr>
<tr>
<td>Glycerine/Glycerol</td>
<td>Glycine</td>
<td>1.293</td>
<td>-2.395</td>
<td>-12.615</td>
<td>KEGG ID</td>
<td>C00037</td>
<td>Glycerine</td>
</tr>
<tr>
<td>3-(indol-3-yl)methacrylic acid</td>
<td></td>
<td>-1.218</td>
<td>0.006</td>
<td>0.297</td>
<td>KEGG ID</td>
<td>C00331</td>
<td>Indoleacrylic acid</td>
</tr>
<tr>
<td>L-2-Aminoadipic acid</td>
<td></td>
<td>-0.006</td>
<td>-0.222</td>
<td>0.276</td>
<td>KEGG ID</td>
<td>C00956</td>
<td>L-2-Aminoadipic acid</td>
</tr>
<tr>
<td>L-formylkynurenine</td>
<td>L-Formylkynurenine</td>
<td>-14.49</td>
<td>-0.041</td>
<td>0.287</td>
<td>CAS Number</td>
<td>1022-31-7</td>
<td>L-Formylkynurenine</td>
</tr>
<tr>
<td>Lysine</td>
<td></td>
<td>-3.412</td>
<td>0.087</td>
<td>-1.116</td>
<td>KEGG ID</td>
<td>C00047</td>
<td>Lysine</td>
</tr>
<tr>
<td>N-Acetyl-L-glutamine</td>
<td>N-Acetyl-L-glutamine</td>
<td>0.342</td>
<td>-0.107</td>
<td>-0.201</td>
<td>KEGG ID</td>
<td>C00624</td>
<td>N-Acetyl-L-glutamine</td>
</tr>
<tr>
<td>N1-acetylserotonin</td>
<td>N1-Acetylserotonin</td>
<td>0.047</td>
<td>-0.253</td>
<td>0.134</td>
<td>KEGG ID</td>
<td>C00612</td>
<td>N1-Acetylserotonin</td>
</tr>
<tr>
<td>Orthophosphate</td>
<td>Phosphoric acid</td>
<td>-0.49</td>
<td>-0.227</td>
<td>0.337</td>
<td>KEGG ID</td>
<td>C00009</td>
<td>Phosphoric acid</td>
</tr>
<tr>
<td>Proline</td>
<td></td>
<td>-0.059</td>
<td>0.023</td>
<td>0.01</td>
<td>KEGG ID</td>
<td>C00148</td>
<td>Proline</td>
</tr>
<tr>
<td>Saccharopine</td>
<td>Saccharopine + ...</td>
<td>14.49</td>
<td>0</td>
<td>0</td>
<td>KEGG ID</td>
<td>C00049</td>
<td>Saccharopine + ...</td>
</tr>
<tr>
<td>Spermine</td>
<td>Spermine</td>
<td>-2.16</td>
<td>0.215</td>
<td>0.222</td>
<td>KEGG ID</td>
<td>C00750</td>
<td>Spermine</td>
</tr>
<tr>
<td>Succinoadenosine</td>
<td>Succinoadenosine</td>
<td>-1.318</td>
<td>-0.039</td>
<td>1.548</td>
<td>CAS Number</td>
<td>4542-23-8</td>
<td>Succinoadenosine</td>
</tr>
<tr>
<td>Urea</td>
<td>Urea</td>
<td>0.07</td>
<td>-0.164</td>
<td>-0.488</td>
<td>KEGG ID</td>
<td>C00086</td>
<td>Urea</td>
</tr>
<tr>
<td>Urocanate</td>
<td>Urocanic acid</td>
<td>0.321</td>
<td>-6.908</td>
<td>-0.047</td>
<td>KEGG ID</td>
<td>C00785</td>
<td>Urocanic acid</td>
</tr>
</tbody>
</table>

Heatmap of all pathway entities, dynamically linked to pathway selection for comparative analysis
Example: Glucose Pathway

Glycolysis

Lipolysis

Ketogenesis

Proteolysis

Insulin

Hippuric acid

Glucose

Hypoxanthine

Bile salts

GCA
GDCA
TCDCA

Table:

<table>
<thead>
<tr>
<th>Compound</th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>-0.873</td>
<td>-0.007</td>
<td>0.281</td>
<td>KEGG ID</td>
</tr>
<tr>
<td>D-Glucose</td>
<td>0.58</td>
<td>-0.02</td>
<td>-0.517</td>
<td>HMP ID</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>-2.611</td>
<td>-0.017</td>
<td>1.478</td>
<td>HMP ID</td>
</tr>
<tr>
<td>Lysine</td>
<td>-3.412</td>
<td>0.087</td>
<td>-1.116</td>
<td>KEGG ID</td>
</tr>
<tr>
<td>Pyruvic acid</td>
<td>-0.785</td>
<td>0.112</td>
<td>0.04</td>
<td>HMP ID</td>
</tr>
</tbody>
</table>
Agilent Metabolomics Solutions

- Most comprehensive metabolomics portfolio
- Robust, sensitive and accurate LC/MS and GC/MS hardware offerings
- Fast and powerful naïve compound feature extraction algorithm (MFE)
- Metabolite identification techniques available for both LC/MS and GC/MS data
- Statistical analysis package built for mass spectrometric chemometric analysis
- Biological pathway analysis package integrated into software workflow
Acknowledgements

AFO Metabolomics

• Sumit Shah
• Daniel Cuthbertson
• Anne Blackwell
• Mark Sartain
• Stephan Bauman

AE EMEA Metabolomics

• Joachim Thiemann
• Friedrich Mandel
• Gordon Ross
• Donna Potts
• Thomas Hennessy (APO)

Santa Clara Metabolomics Group

Theo Sana, Steve Fischer, Norton Kitagawa