The ENTACT Story: Using US EPA Resources to Evaluate and Enhance Non-target Workflows

Jon Sobus\textsuperscript{1}, Elin Ulrich\textsuperscript{1}, Jarod Grossman\textsuperscript{2}, Alex Chao\textsuperscript{2}, Seth Newton\textsuperscript{1}, Antony Williams\textsuperscript{1}, Ann Richard\textsuperscript{1}, Chris Grulke\textsuperscript{1}, Andrew McEachran\textsuperscript{2}, Randolph Singh\textsuperscript{2}, Hussein Al-Ghoul\textsuperscript{2}

\textsuperscript{1} Center for Computational Toxicology and Exposure
\textsuperscript{2} ORAU/ORISE Participant
1) Research drivers for NTA/NTS
2) ENTACT genesis & study design
3) Progress to date (at EPA)
4) Progress to date (outside EPA)
5) Ongoing & future work
Many industrial & commercial chemicals are covered by the Toxic Substances Control Act (TSCA), which is administered by EPA.

TSCA updated in June 2016 to allow risk-based evaluation of existing and new chemicals.

Characterization of risk requires exposure and hazard data.

EPA’s Office of Research and Development (ORD) is developing new approach methodologies (NAMs) for rapid risk characterization.

NTA is a promising NAM, but requires careful evaluation and implementation.
“The novelty of nontarget analysis, particularly its current lack of implementation by regulatory agencies, has prevented the establishment of streamlined quality assurance and quality control (QA/QC) procedures.”

“No single analytical technique is suitable for the analysis of all compounds, and successful nontargeted screening will require the development of multiplatform approaches, facilitated and validated through interlaboratory collaborations.”
Key Research Needs

- 18 Institutes
- 12 Countries
- 1 river water extract

- Workflows & Methods:
  - Analytical → well harmonized
  - Data processing → not harmonized

Clearly expressed needs for:

1) More tightly defined interlaboratory comparisons
2) The use of spiked samples
3) The shared use of comprehensive suspect lists
EPA/ORD Takes a Leadership Role

Non-Targeted Analysis Workshop

The U.S. Environmental Protection Agency (EPA) will host the Non-Targeted Analysis Workshop August 18-19, 2015 at EPA's Research Triangle Park Campus.

EPA's ENTACT Study Breaks New Ground with Non-Targeted Research

Published July 30, 2018

EPA scientists are leading a multi-phase project to evaluate the ability of non-targeted analysis laboratory methods to consistently and correctly identify unknown chemicals in samples. EPA's Non-Targeted Analysis Collaborative Trial (ENTACT) was formed in late 2015 and includes nearly 30 academic, government, and industry groups. Non-targeted analysis involves analyzing water, soil and other types of samples to identify unknown chemicals that may be present, without having a preconceived idea of what chemicals may be in the samples.

“One of our main goals is to find out what scientists are doing with non-targeted analysis as a group at large, particularly which chemicals we correctly identify and why,” says Elm Ulrich, an EPA scientist who co-leads ENTACT with EPA's Jon Sobus.
Science Questions for Research Community

- How variable are tools and results from lab to lab?
- Are some methods/tools better than others?
- How does sample complexity affect performance?
- What chemical space does a given method cover?
- How sensitive are specific instruments/methods?

EPA’s Non-Targeted Analysis Collaborative Trial
Original ENTACT Concept

ToxCast Chemicals

Lab A

Why are certain chemicals only found with certain methods?

Lab B measurement space

Lab A measurement space

Lab C measurement space

What impurities/interaction products found?

Can we expand coverage?

Can we model these behaviors?

? “other” space (missing chemicals)
**ENTACT Part 1**

*Chemicals from ToxCast Library*

~1200 ToxCast Chemicals  
(highest quality)  

10 Mixtures  
(100-400 chemicals each)  

Multi-Well Plates*  

- Reference & Fortified House Dust  
- Reference & Fortified Human Serum  
- Reference & Fortified Silicone Wristbands

**ENTACT Part 2**

~25 Collaborators & 5 Contractors*:

1\textsuperscript{st}: Blinded analysis  
2\textsuperscript{nd}: Unveiling of chemicals  
3\textsuperscript{rd}: Unblinded evaluation
Design of ENTACT Mixtures

Number of Chemicals

Mixture Number

- 5 NTA method replicates
- Grade A - replicate 90 set
- Grade A - unique to mix
- Grade A - all isobaric set (replicated)
- Grades B,C - lower purity mix
Resources Provided to Participants

- SOPs for sample handling, analysis, and data submission
- Procedures used for sample preparation
- Up to 16 samples with eventual (unblinded) chemical mappings
- MS-Ready DSSTox list (671,852 unique) with .mol files
- MS-Ready ToxCast list (4,248 unique) with .mol files
- Method and Data reporting templates
- FTP site, accounts, and instructions
EPA Methods for ENTACT Mixtures

Agilent ZORBAX Eclipse Plus C8 column (2.1 x 50 mm, 1.8 μm)
A: 5% methanol, 95% water (0.4 mM ammonium formate)
B: 95% methanol, 5% water (0.4 mM ammonium formate)

Waters Acquity UPLC® BEH C18 column (2.1 x 50 mm, 1.7 μm)
A: water (0.1% formic acid)
B: acetonitrile (0.1% formic acid)

10 ENTACT Mixtures

Agilent 6530B Q-TOF

3 ENTACT Mixtures
**Sample Preparation**
3 Dilutions, 3 Replicate Injections, 6 Blanks

**Sample Analysis**
LC-QTOF/HRMS: ESI+ and ESI-, MS1

**Peak Picking & Alignment**
Agilent MassHunter Profinder Software

**Formula Assignment**
Agilent Mass Profiler Professional Software, DSSTox Unique MS-Ready Formula List

**Feature Filtering & Flagging**
Custom Script:
Blank Subtraction, Fold-Change Thresholds, Formula Match Score Cut-Off

**Candidate Structure Selection**
EPA CompTox Chemicals Dashboard:
Batch Search, Data Source Ranking

**Chemical Structure Corroboration**
LC-QTOF/HRMS:
DDA MS2 Using Preferred Ions List, Agilent Reference MS2 libraries

---

**Identifying known unknowns using the US EPA’s CompTox Chemistry Dashboard**
Andrew D. McEachran1 • Jon R. Sobus2 • Antony J. Williams3

---

**“MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies**
Andrew D. McEachran1,2, Karel Mansoun1,2•, Chris Gruifke2, Emma L. Schymanski2, Christoph Ruttkies2 and Antony J. Williams2

---

**Journal of Cheminformatics**
Open Access

---

DOI 10.1016/j.aca.2017.06.0216
Rapid Communication
EPA Initial Results

By Feature (total = 26K)

- Artifacts: 44%
- Unmatched Features: 52%
- Matched Features: 4%

~ 25% of Observed Features Matched to a Spiked Substance

By Substance (total = 1,269)

- Not Observed: 49%
- Limited Evidence: 14%
- Observed (MS1): 11%
- Observed (MS2): 26%

~ 75% of Spiked Substances were Observed

* Only 48% of ENTACT substances were in reference MS² library
**Generation of in silico Spectra**

**CFM-ID v2.0**

- Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification

**Authors**: Andrew D. McEachran, Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghouli, Chris Gruke, Jon R. Sobus & Antony J. Williams

**Training Set**: Metlin MS2 spectra and structures

**Machine Learning**

**Fragmentation Prediction Model**

- DSSTox MS-Ready Structures (~765,000)
- DSSTox MS2 spectra (10, 20, 40v)

**Linking in silico MS/MS spectra with chemistry data to improve identification of unknowns**


Reference vs. *in silico* Library Coverage

<table>
<thead>
<tr>
<th>MS2 Library</th>
<th>% of “Pass” Compounds Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agilent PCDL</td>
<td>53%</td>
</tr>
<tr>
<td>CFM-ID Top Hit</td>
<td>50%</td>
</tr>
<tr>
<td>PCDL and/or CFM-ID Top Hit</td>
<td>73%</td>
</tr>
</tbody>
</table>

PCDL → Agilent reference MS² library

“Pass” compounds (n=377) → ENTACT chemicals observed with MS² data
Who Else is Working on ENTACT?

**Contractors:**
- University of Alberta
- Emory University
- Duke University
- Oregon State University
- San Diego State University
- Pacific Northwest National Laboratory

**Vendors:**
- Agilent Technologies
- Thermo Scientific

**General Participants:**
- 19 Blind submissions
- 15 Unblinded submissions
# Comparing Reported Features (n=16 labs)

<table>
<thead>
<tr>
<th>Act.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Dust</th>
<th>Serum</th>
<th>Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>144</td>
<td>166</td>
<td>187</td>
<td>292</td>
<td>269</td>
<td>318</td>
<td>470</td>
<td>177</td>
<td>410</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>2</td>
<td>142</td>
<td>154</td>
<td>102</td>
<td>129</td>
<td>250</td>
<td>242</td>
<td>401</td>
<td>399</td>
<td>105</td>
<td>452</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>40</td>
<td>48</td>
<td>59</td>
<td>110</td>
<td>101</td>
<td>97</td>
<td>130</td>
<td>37</td>
<td>109</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>4</td>
<td>301</td>
<td>130</td>
<td>375</td>
<td>341</td>
<td>408</td>
<td>404</td>
<td>719</td>
<td>687</td>
<td>198</td>
<td>327</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>66</td>
<td>74</td>
<td>72</td>
<td>105</td>
<td>118</td>
<td>193</td>
<td>215</td>
<td>54</td>
<td>162</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>6</td>
<td>587</td>
<td>552</td>
<td>596</td>
<td>554</td>
<td>798</td>
<td>846</td>
<td>1327</td>
<td>1274</td>
<td>509</td>
<td>1176</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
<td>114</td>
<td>116</td>
<td>106</td>
<td>182</td>
<td>201</td>
<td>360</td>
<td>374</td>
<td>73</td>
<td>330</td>
<td>236</td>
<td>92</td>
<td>124</td>
</tr>
<tr>
<td>8</td>
<td>337</td>
<td>372</td>
<td>303</td>
<td>365</td>
<td>321</td>
<td>363</td>
<td>466</td>
<td>505</td>
<td>510</td>
<td>463</td>
<td>259</td>
<td>222</td>
<td>313</td>
</tr>
<tr>
<td>9</td>
<td>135</td>
<td>130</td>
<td>125</td>
<td>154</td>
<td>188</td>
<td>195</td>
<td>284</td>
<td>295</td>
<td>100</td>
<td>153</td>
<td>270</td>
<td>54</td>
<td>101</td>
</tr>
<tr>
<td>10</td>
<td>70</td>
<td>57</td>
<td>64</td>
<td>66</td>
<td>105</td>
<td>115</td>
<td>176</td>
<td>125</td>
<td>35</td>
<td>159</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>11a</td>
<td>595</td>
<td>486</td>
<td>571</td>
<td>630</td>
<td>746</td>
<td>669</td>
<td>899</td>
<td>910</td>
<td>588</td>
<td>792</td>
<td>1009</td>
<td>614</td>
<td>NR</td>
</tr>
<tr>
<td>11b</td>
<td>66</td>
<td>170</td>
<td>51</td>
<td>41</td>
<td>272</td>
<td>116</td>
<td>214</td>
<td>101</td>
<td>163</td>
<td>404</td>
<td>861</td>
<td>145</td>
<td>557</td>
</tr>
<tr>
<td>12</td>
<td>51</td>
<td>37</td>
<td>35</td>
<td>39</td>
<td>74</td>
<td>59</td>
<td>124</td>
<td>109</td>
<td>42</td>
<td>105</td>
<td>124</td>
<td>52</td>
<td>76</td>
</tr>
<tr>
<td>13</td>
<td>137</td>
<td>65</td>
<td>45</td>
<td>74</td>
<td>68</td>
<td>234</td>
<td>413</td>
<td>408</td>
<td>120</td>
<td>317</td>
<td>389</td>
<td>178</td>
<td>88</td>
</tr>
<tr>
<td>14</td>
<td>215</td>
<td>249</td>
<td>212</td>
<td>249</td>
<td>207</td>
<td>275</td>
<td>245</td>
<td>254</td>
<td>140</td>
<td>253</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>15</td>
<td>1298</td>
<td>1258</td>
<td>1304</td>
<td>1209</td>
<td>1651</td>
<td>1641</td>
<td>2520</td>
<td>2538</td>
<td>1202</td>
<td>2193</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>16</td>
<td>153</td>
<td>217</td>
<td>221</td>
<td>199</td>
<td>254</td>
<td>321</td>
<td>523</td>
<td>651</td>
<td>496</td>
<td>396</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

- **Act.** = actual measurements

**Color Coding**
- **Blue** = under reported
- **Green** = near actual
- **Red** = over reported
- **NR** = not reported

Comparing Identified Compounds (n=3 labs)

1,269 Spiked Substances

GC = gas chromatography

ESI- = neg. electrospray ionization (liquid chromatography)

ESI+ = pos. electrospray ionization (liquid chromatography)

Experiments with SRM Dust

1) Solvent spike (best case)

2) Post-extraction high spike

3) Pre-extraction high spike

4) Pre-extraction low spike (ENTACT sample)

Newton et al. [in preparation]
Experiments with SRM Dust

Results for Unfortified SRM Dust

<table>
<thead>
<tr>
<th>Chemical Class</th>
<th>All Reported Compounds</th>
<th>Reported Using LC-ESI</th>
<th>Observed Using NTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAHs</td>
<td>69</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PCBs</td>
<td>44</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PFAS</td>
<td>31</td>
<td>31</td>
<td>12</td>
</tr>
<tr>
<td>BFRs</td>
<td>30</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>OCPs</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OPEs</td>
<td>12</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Phthalates</td>
<td>7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>208</td>
<td>43</td>
<td>18</td>
</tr>
</tbody>
</table>

- **Reported using LC-ESI**
- **Not observed using NTA**
- **Observed using NTA**
- **Spiked in ENTACT samples**
- **Spiked at higher conc.**
- **Est. to cause bioactivity in children**

* "..the dose that would be needed in the most-sensitive 5% of the population to produce a steady-state plasma concentration equal to [the 10th] percentile of the ToxCast AC50 distribution across assays for the given chemical."

Newton et al. [in preparation]

Ring et al.: https://doi.org/10.1016/j.envint.2017.06.004
Publications to date

**Analytical and Bioanalytical Chemistry** (2019) 411:853–866
https://doi.org/10.1002/abdc.20184255

**RESEARCH PAPER**

EPA’s non-targetted analysis collaborative trial (ENTACT): genesis, design, and initial findings

Elin M. Ulrich1 • Jon R. Sobus1 • Christopher M. Grulke2 • Ann M. Richard3 • Seth R. Newton1 • Mark J. Strynar1 • Kamel Mansouri1,4,6 • Antony J. Williams2

Received: 30 July 2018 / Revised: 14 September 2018 / Accepted: 12 October 2018 / Published online: 6 December 2018

© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2018

**Analytical and Bioanalytical Chemistry** (2019) 411:835–851
https://doi.org/10.1002/abdc.201841526

**RESEARCH PAPER**

Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance

Jon R. Sobus1, *1 Jared N. Grossman2, 3, 4 Alex Chao2, 5 Randolph Singh2, 5 Antony J. Williams8, 9 Christopher M. Grulke3, 5, 4, 6 Ann M. Richard3, 5, 9 Seth R. Newton1, 6 Andrew D. McEachran2, 5, 6, 9 Elin M. Ulrich1

Received: 19 September 2018 / Revised: 14 November 2018 / Accepted: 27 November 2018 / Published online: 5 January 2019

© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2019

**Chromatography Today**

February / March 2018

Comprehensive, Non-Target Characterisation of Blinded Environmental Exposure Standards Using GC×GC and High Resolution Time-of-Flight Mass Spectrometry

by Elin M. Ulrich, Todd Richards and Joe Birkley

L.E.C.O., Saint Joseph, Michigan, USA

Corresponding Author: elin_ful@leco.com

**JCIIM**

JOURNAL OF CHEMICAL INFORMATION AND MODELING

published online 5/5/2019

Evaluation of In Silico Multifeature Libraries for Providing Evidence for the Presence of Small Molecules in Synthetic Blinded Samples

Jamie R. Nuñez, 1, 6 Sean M. Colby, 1, 6 Dennis G. Thomas, 1, 6 Malak M. Tlaify, 1, 2 Nikola Tolic, 1, 6 Elin M. Ulrich, 1, 6 Jon R. Sobus, 2, 5 Thomas O. Metz, 6, 5 Justin G. Teegarden, 1, 2, 8 and Ryan S. Rendlow8, 5

1Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99354, United States
2U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, North Carolina 27711, United States
3Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon 97331, United States
4Department of Environmental Science, University of Arizona, Tucson 85712, United States

*Corresponding Author: jnunez@pnnl.gov
Summary of ENTACT Findings

- NTA methods are suitable for detecting many ToxCast chemicals
- False positives can greatly outweigh true positives
  - False Pos / True Pos ~ 10×
  - Work needed on feature credentialing
- True Positives: ≤75%
  - Will miss some chemicals that are present in samples
    - Why? Which ones? Always?
- Multiple methods required for broad characterization
  - No “one size fits all” method
  - Subtle method changes affect measurable chemical space
- Concentration, media, and extraction techniques will affect performance
- Goal reached when we can make these statements:
  - “When a compound is observed, we’re confident it’s really there!”
  - “When a compound isn’t observed, we’re confident it’s not there!”
Ongoing and Future Work

- Full cross-lab performance evaluation
  - Primary focus → true positives, false negatives, confidence levels
  - Secondary focus → unexpected true positives

- Database development
  - Enable user queries, additional analyses, model development

- Global summary report
  - Provide guidance and acceptance criteria for NTA studies

- The benefits of ENTACT will be proportional to the level of effort!
This work was supported, in part, by ORD's Pathfinder Innovation Program (PIP) and an ORD EMVL award.

**EPA ORD**
- Hussein Al-Ghoul*
- Alex Chao*
- Jarod Grossman*
- Kristin Isaacs
- Sarah Laughlin*
- Charles Lowe
- James McCord
- Jeff Minucci
- Seth Newton
- Katherine Phillips
- Tom Purucker
- Randolph Singh*
- Mark Strynar
- Elin Ulrich

* = ORISE/ORAU

**EPA ORD (cont.)**
- Chris Grulke
- Kamel Mansouri*
- Andrew McEachran*
- Ann Richard
- John Wambaugh
- Antony Williams

**Agilent**
- Jarod Grossman
- Andrew McEachran

**GDIT**
- Ilya Balabin
- Tom Transue
- Tommy Cathey
Questions?

sobus.jon@epa.gov

The views expressed in this presentation are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.