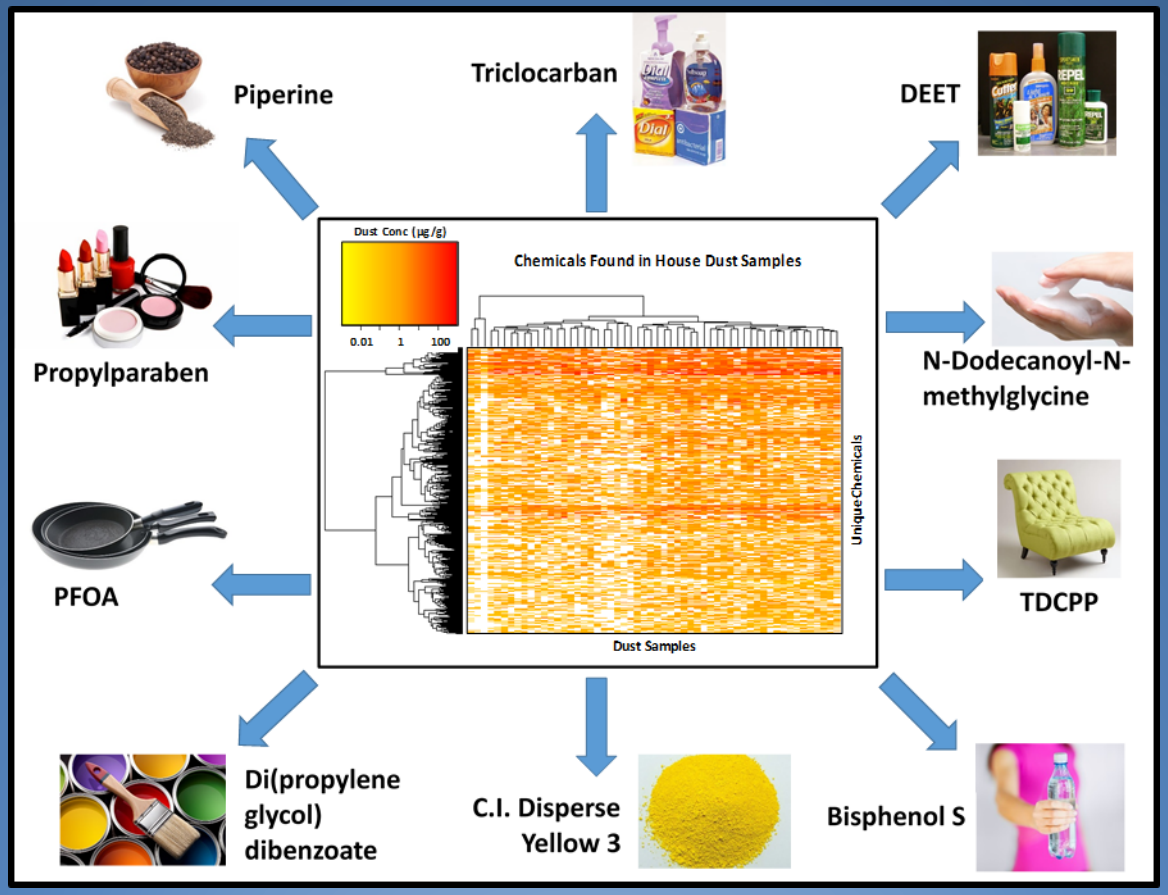
Advancing and Integrating Non-Targeted Analysis Research at the U.S. EPA



Jon R. Sobus
U.S. EPA Office of Research and Development
July 20, 2017

Exposure Science in the 21st Century: What are Key Drivers?

1) Understanding causes of disease

"…70-90% of disease risks are probably due to differences in environments"

EPIDEMIOLOGY

Environment and Disease Risks

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

Stephen M. Rappaport and Martyn T. Smith

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (*1*–*3*). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize "environmental exposures." This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (*4*). A more comprehensive and quantitative view of environmental exposure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of "environmental exposures." In fact, all of these exposure categories can contribute to chronic diseases and should be investigated collectively rather than separately.

To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxicants) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (*5*, *6*) (see the figure). This internal chemical environment continually fluctuates during life due to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

The term "exposome" refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

School of Public Health, University of California, Berkeley, CA 94720–7356, USA. E-mail: srappaport@berkeley.edu

Published by AAAS

2) Ensuring chemical safety and human/eco health

GIVE A DOG A PHONE
Technology for our furry friends

NewScientist

We've made 150,000 new chemicals

We touch them, we wear them, we eat them

But which ones should we worry about?

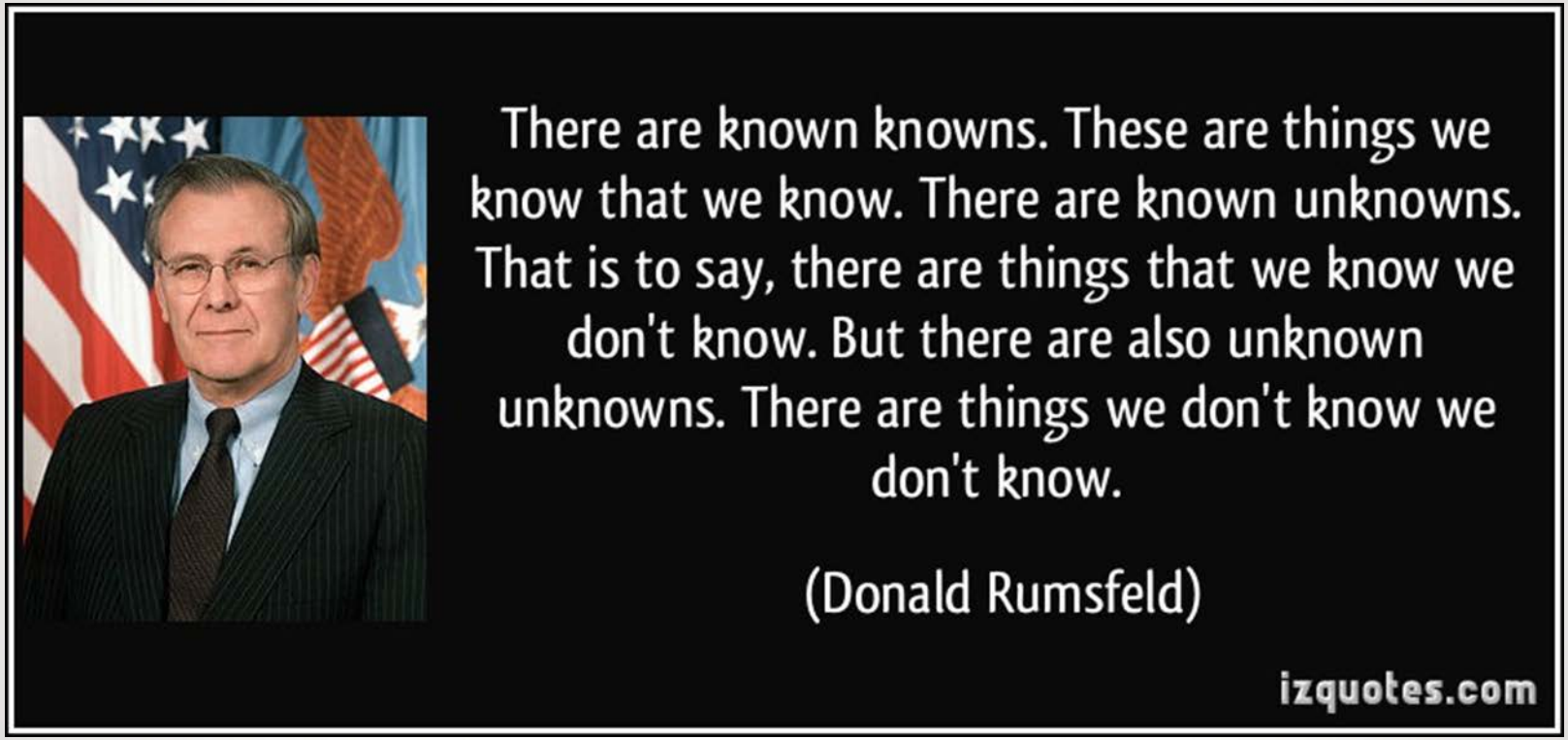
SPECIAL REPORT, page 34

THE GOOD FIGHT
Most violence is also virtuous

CHAMBER OF SECRETS
The greatest ever find of early human bones

IS IT ALIVE?
Artificial worm could be first digital animal

How Much Do We Know About Exposure?



Analytical Chemistry Jargon

Targeted Analysis

- The “known knowns”
  - <<1% of the exposome

We can’t solve 21st century public health problems if we’re blinded to +99% of exposure data

Suspect Screening Analysis (SSA)

- The “known unknowns”
  - ~5-10% of the exposome

We need efficient methods capable of rapidly measuring poorly studied compounds

Non-Targeted Analysis (NTA)

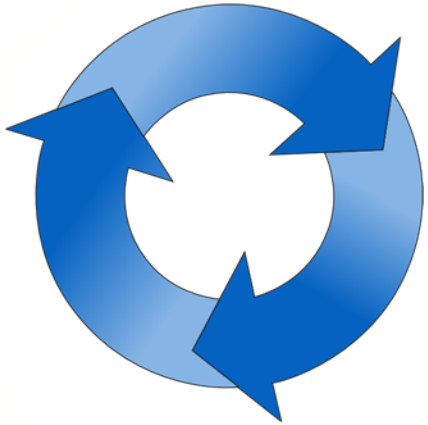
- The “unknown unknowns”
  - 90-95% of the exposome

We need ways to characterize compounds of interest that aren’t yet known to exist

We often use “NTA” as a generic term for any research that isn’t “targeted”

NTA: Tools of the Trade

Analytical Instruments



Comp. Tools & Workflows

FOR IDENT

Chemical Databases

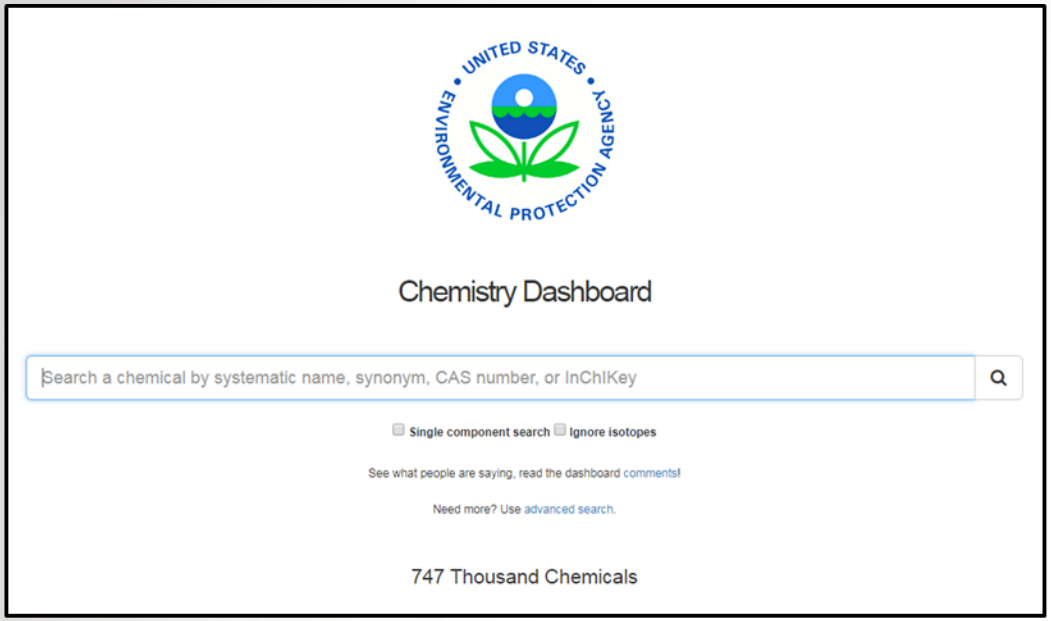
MassBank

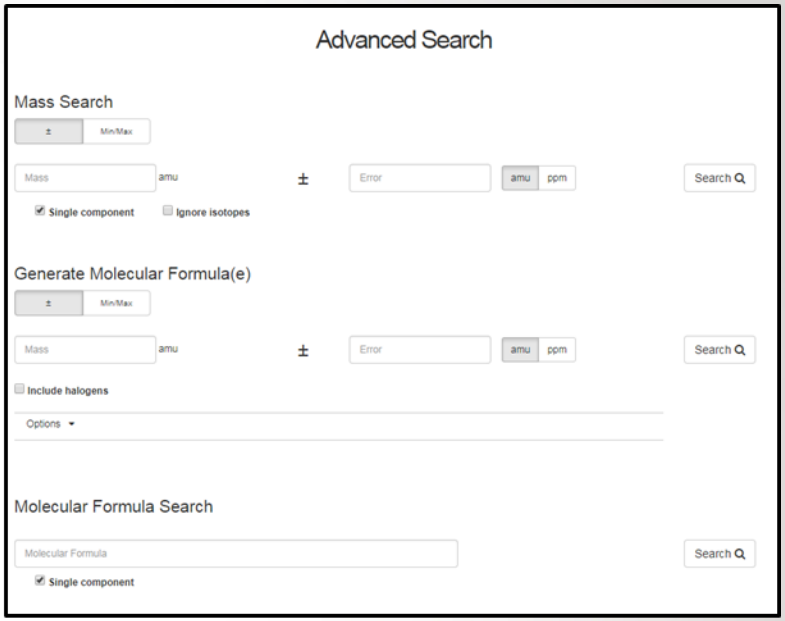
PubChem

ChemSpider
The free chemical database

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Chemistry Dashboard

EPA CompTox Chemistry Dashboard



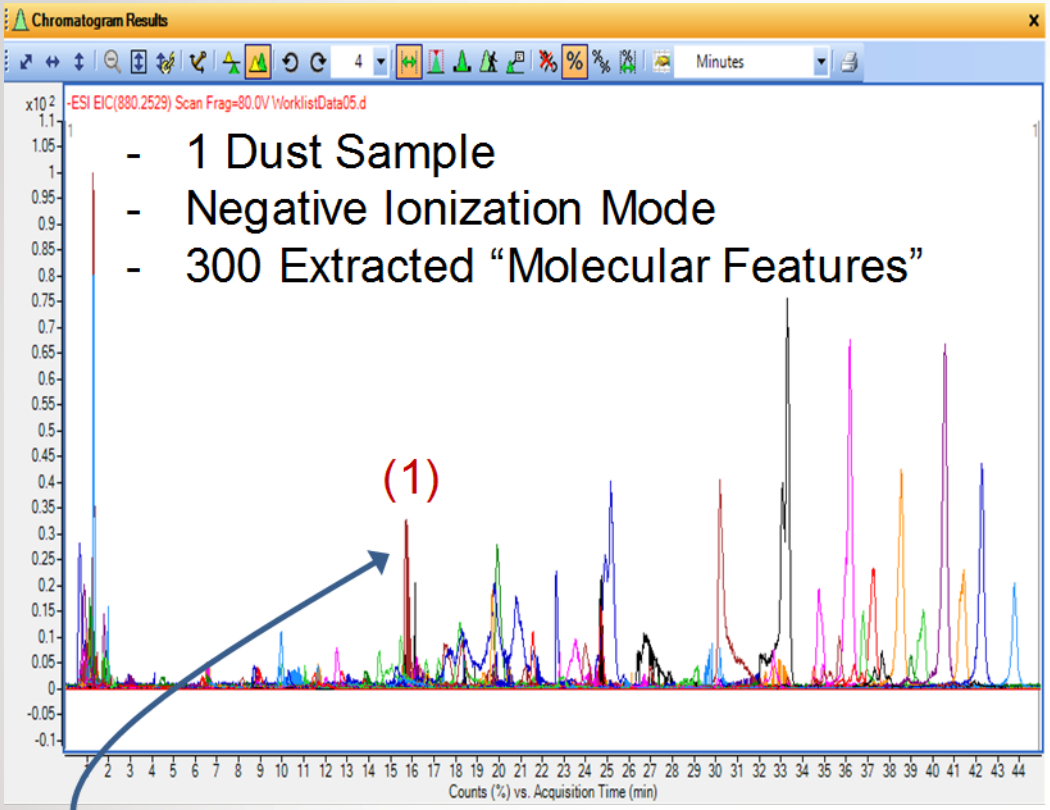


https://comptox.epa.gov/dashboard

- Chemical Properties
- Structural Identifiers
- Exposure & Use Data
- Tox. & Bioactivity Data
- Much more…

Project Lead: Antony Williams,
williams.antony@epa.gov

General Goals of NTA



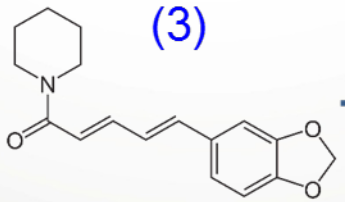
1) Prioritize "molecular features"

2) Correctly assign formulas

3) Correctly assign structures

4) Predict chemical concentrations

5) Determine chemical sources

(2) $C_{17}H_{19}NO_3$ → (3)  → (4) 12 µg/g → (5) 

Challenges of NTA

- Difficulty handling all observed features
  - Data processing and data storage limits
- Analytical focus on features that stand out
  - Bias towards high-intensity compounds
  - Bias towards compounds that are enriched
- No standardized methods
  - Difficult to compare results across labs
- Conventional confirmation uses analytical standards
  - Many chemicals not previously studied → no standards
  - Thousands of chemicals to confirm
  - EXTREMELY low throughput

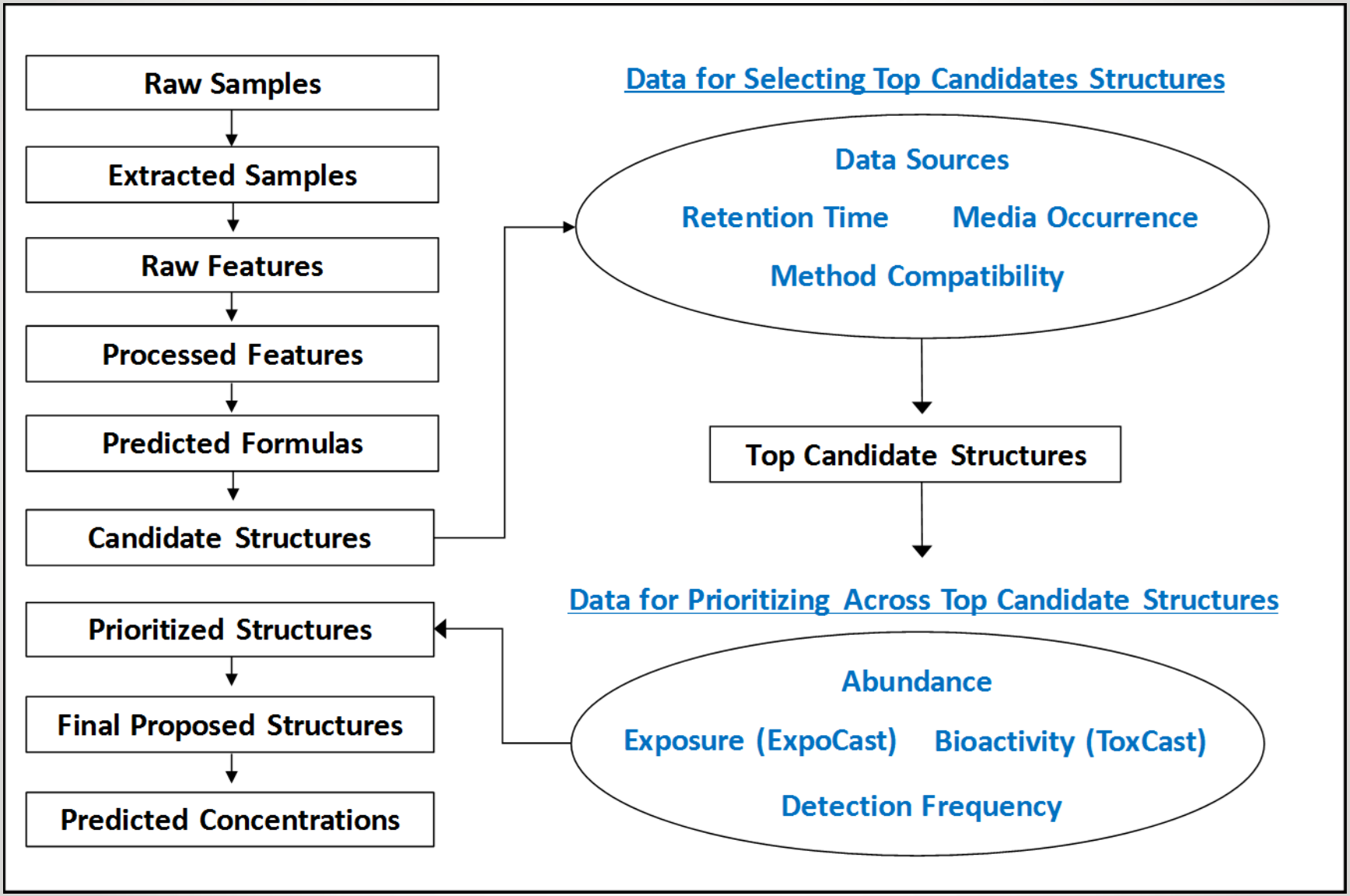
Challenges of NTA

- Difficulty handling all observed features
  - Data processing and data storage limits
- Analytical foc
  - Bias toward
  - Bias towards compounds that are enriched
- No standardized methods
  - Difficult to compare results across labs
- Conventional tandards
  - Many chem standards
  - Thousands of chemicals to confirm
  - EXTREMELY low throughput

New NTA Workflows and Analysis Tools

EPA-led Collaborative NTA Trial

EPA NTA Workflow



EPA NTA Workflow

- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- Predicted Formulas
- Candidate Structures
- Prioritized Structures
- Final Proposed Structures
- Predicted Concentrations





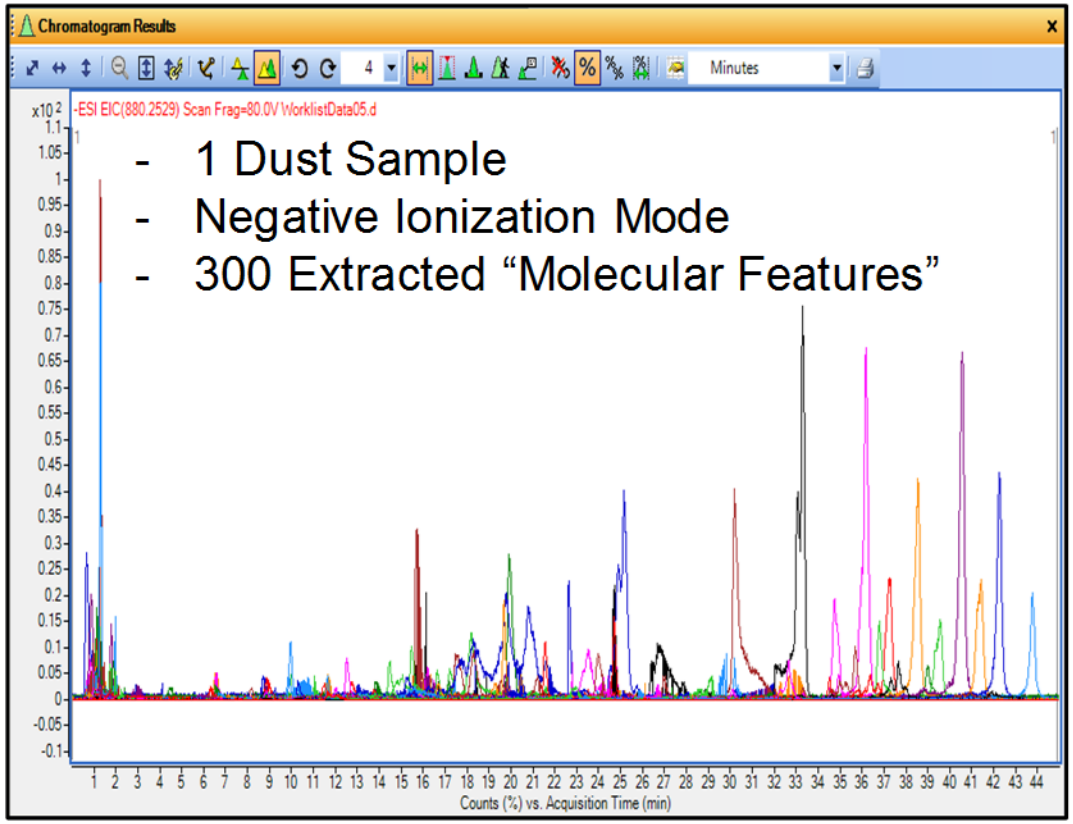




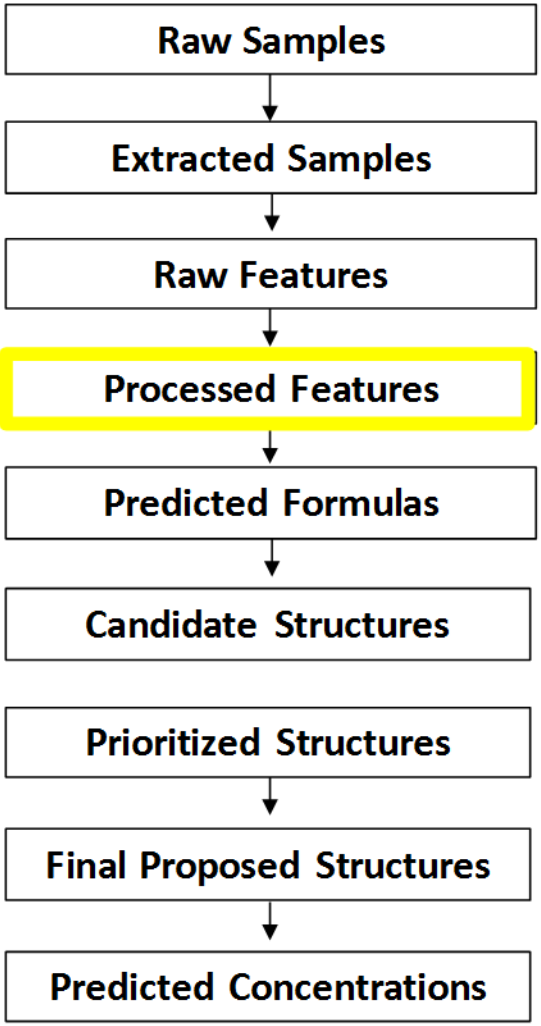


EPA NTA Workflow

- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- Predicted Formulas
- Candidate Structures
- Prioritized Structures
- Final Proposed Structures
- Predicted Concentrations



EPA NTA Workflow



	C	D	E	F	G	P	Q	R	U	V	W
1	Ionization_Mode	Compound	Score	Mass	Retention_Time	Sample_AE_1	Sample_AE_2	Sample_AE_3	Sample_BE_1	Sample_BE_2	Sample_BE_3
2	Esi+	C5H9NO	87.6	99.0684	0.547	893735	900226	912425	883082	898472	925539
3	Esi+	C6H12O	79.9	100.089	2.720	313654	322144	324694	284466	293754	303573
4	Esi+	C5H11NO	78.66	101.0839	0.928	282687	548672	277927	518837	536016	536266
5	Esi+	C4H10O3	99.88	106.0631	2.720	1150011	1157381	1185995	1004359	1055811	1067739
6	Esi+	C4H10O3	87.72	106.0632	13.864	867270	878146	859412	860052	880647	884363
7	Esi+	C4H10O3	86.19	106.0632	14.561	381606	367729	371812	351883	335371	340179
8	Esi+	C7H8O	82.76	108.0577	4.179	82154	92712	88446	69079	74237	75031
9	Esi+	C16H24	68.21	108.0935	17.564	134690	134577	135563	132266	132708	131244
10	Esi+	C8H12	84.83	108.0939	19.470	152306	131197	129558	142948	131532	128163
11	Esi+	C6H7NO	76.52	109.0523	0.913	135066	134029	134545	145387	143031	150459
12	Esi+	110.0399@0.5809999		110.0399	0.581	76841	78264	71477	71198	77970	80148
13	Esi+	110.0832@1.094		110.0832	1.094	82736	84607	83462	69141	70535	74819
14	Esi+	C6H10N2	78.6	110.0842	0.780	104967	93167	94364	76166	75748	73172
15	Esi+	C8H14	83.61	110.109	12.779	152534	154176	151465	149541	151543	162820
16	Esi+	C8H14	82.9	110.1097	11.784	153131	131030	146960	88310	91205	82244
17	Esi+	C8H16	77.5	112.1249	8.159	105224	112921	110816	78584	97387	101468
18	Esi+	C8H16	86.93	112.125	13.947	273088	260100	249562	283827	297367	277256
19	Esi+	C8H16	87.43	112.1251	15.689	340461	322983	353968	316599	323923	344990
20	Esi+	C8H16	87.9	112.1253	16.186	1341392	1278715	1286083	1280351	1301173	1309748
21	Esi+	C8H16	87.67	112.1254	16.402	9757336	9862888	9864919	9548312	9528605	9578356
22	Esi+	C6H11NO	78.18	113.0839	1.095				239846	255280	
23	Esi+	C6H11NO	91.77	113.084	0.829	6534311	6566026	6452624	6179379	6531283	6613715
24	Esi+	C6H10O2	67.89	114.0684	10.796	47861	48356	51116	56268	48446	
25	Esi+	C6H13NO	87.04	115.0998	1.842	793262	813696	805649	814047	839653	857749
26	Esi+	C4H4O4	87.17	116.0111	15.499	81179	76490	80131	80882	78465	79062
27	Esi+	C6H12O2	79.36	116.0835	0.664	846477	846464	846044	797082	813503	816182
28	Esi+	C6H12O2	69.99	116.0836	5.755	392157	393872	378580	340458	342060	368264
29	Esi+	C6H12O2	79.64	116.0837	5.888	1025006	1035319	1054898	913993	926335	947337
30	Esi+	C5H11NO2	81.14	117.0793	0.516	124358	134061	121452	108516	130971	138120
31	Esi+	C6H15NO	85.68	117.1147	0.539	182748	170327	189521	168564	179837	182100
32	Esi+	C6H15NO	79.93	117.1148	0.539	201399	95853	169240	180017	184911	174209

EPA NTA Workflow

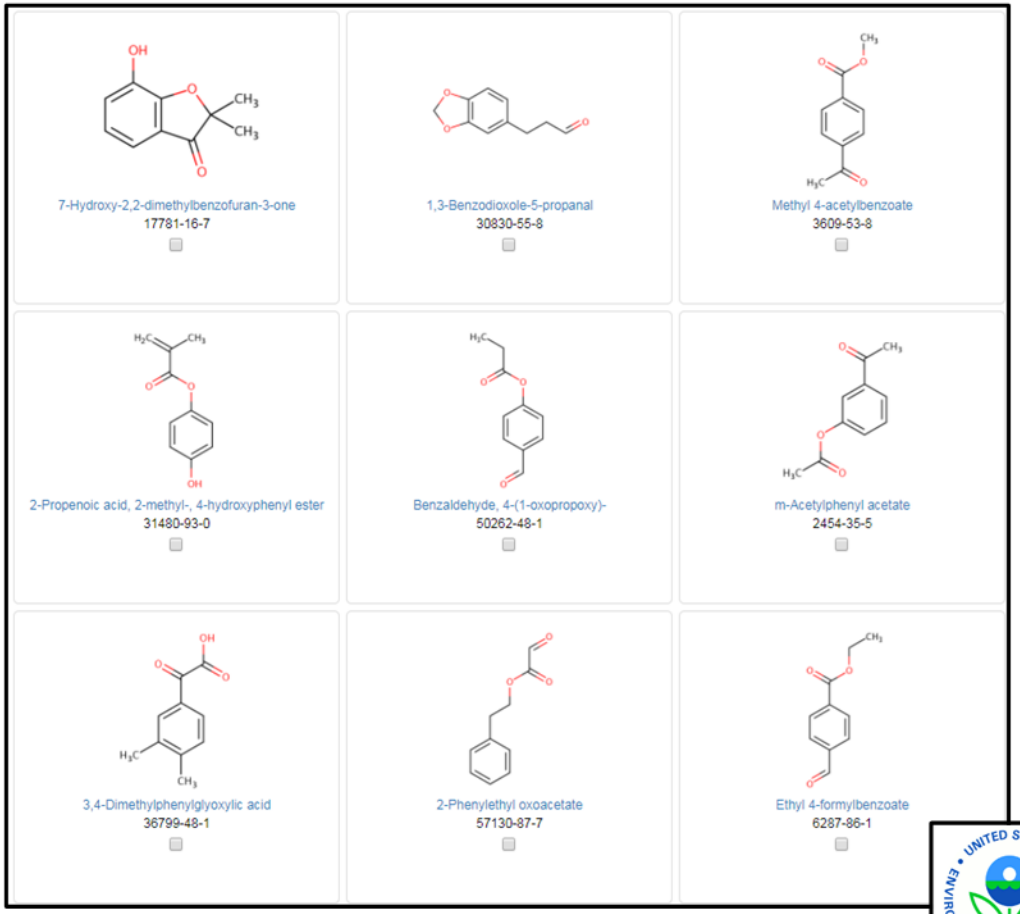
- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- **Predicted Formulas**
- Candidate Structures
- Prioritized Structures
- Final Proposed Structures
- Predicted Concentrations

Compound	For_Dashboard_Search
C10BrF21	1
C10Cl10O	1
C10Cl2F20	1
C10H10BrClO2S	1
C10H10O2	1
C10H10O3	1
C10H12N6O5	1
C10H12O3	1
C10H12O4	1
C10H14N4O10P2S	1
C10H15N2P	1
C10H15NO	1
C10H15O4P	1
C10H15OP	1
C10H16N4O17P4	1
C10H16N5O11P3	1
C10H17O3P	1
C10H19NO5PS	1
C10H20Cl4O8P2	1
C10H20O2	1
C10H21NO2	1
C10H21O2P	1
C10H22O6	1
C10H23NO	1
C10H27N4P	1
C10H2F18O2	1
C10H3Cl3F3N3O2	1
C10H3F15O	1
C10H4Cl6N2O2S2	1
C10H4Cl6N2O6	1
C10H4Cl6S2	1
C10H5Br4NO2	1
C10H5BrCl3NO2S2	1
C10H5BrClNO2S	1
C10H5F17O3S	1
C10H6Br2Cl6	1
C10H6Br4O4	1

https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

EPA NTA Workflow

- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- Predicted Formulas
- **Candidate Structures**
- Prioritized Structures
- Final Proposed Structures
- Predicted Concentrations



Chemistry Dashboard

EPA NTA Workflow

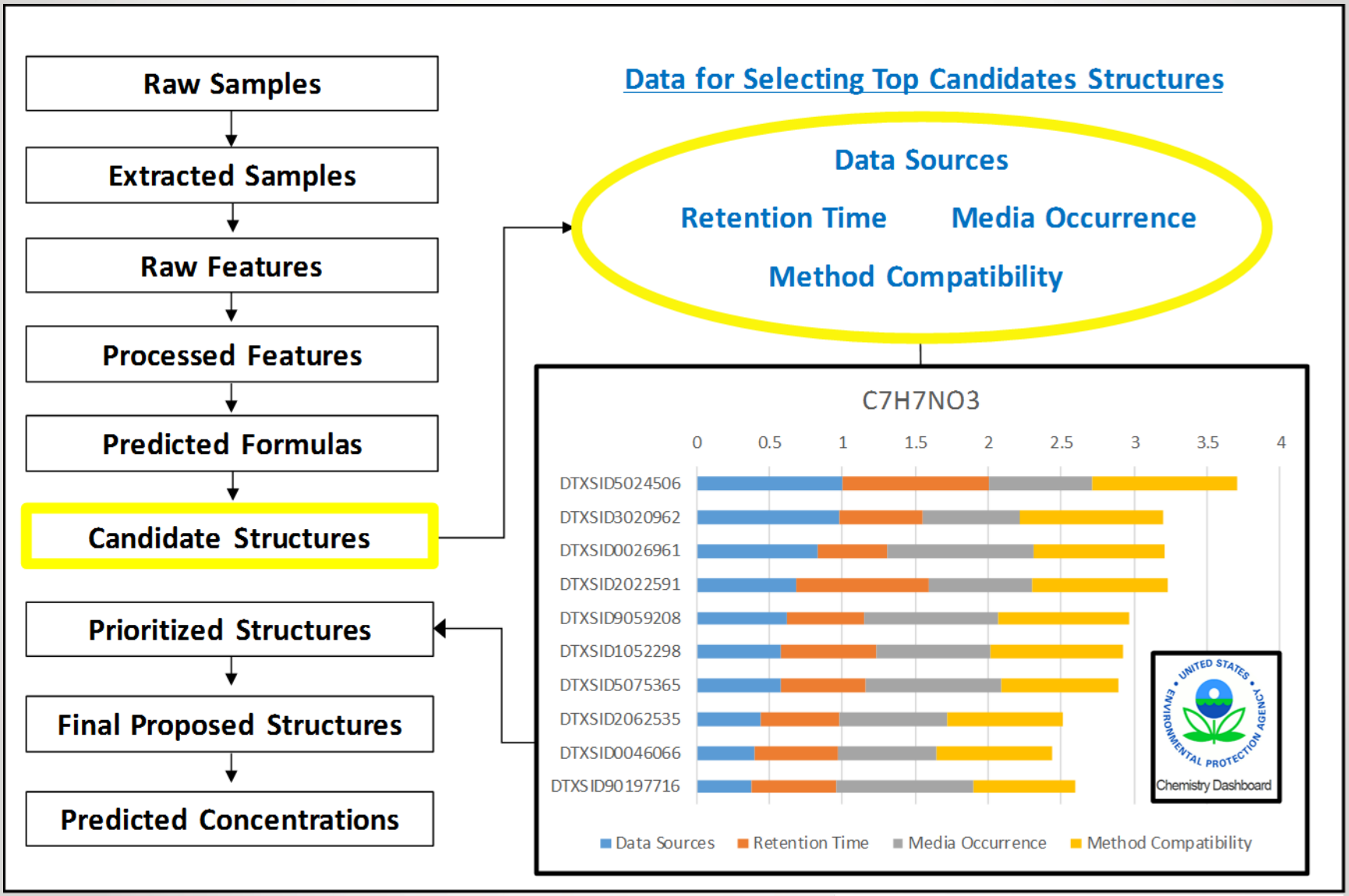


Figure from Andrew McEachran

EPA NTA Workflow

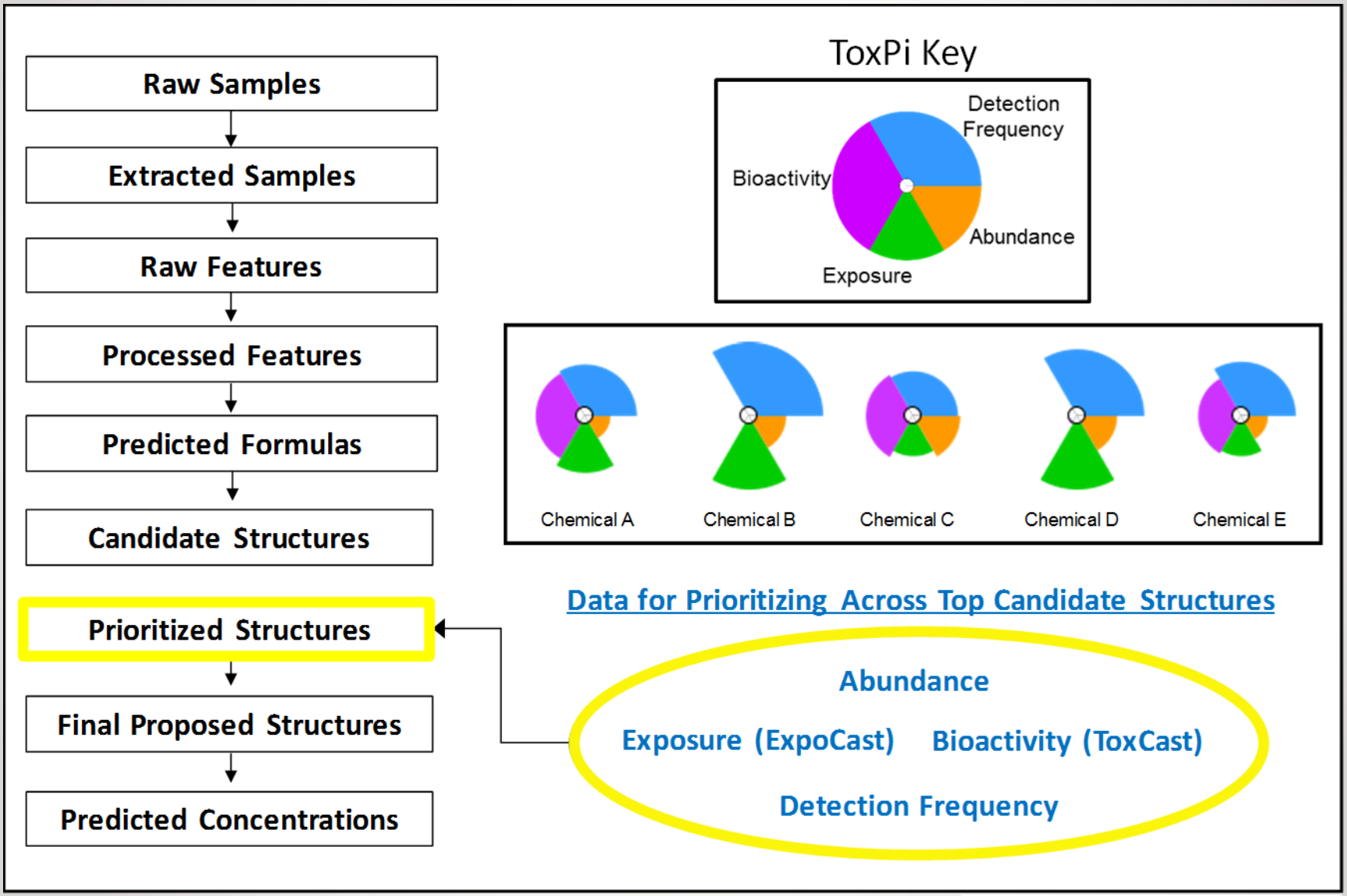
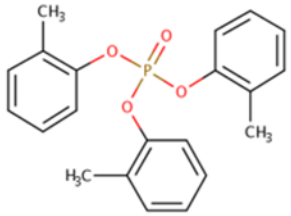


Figure from Julia Rager

EPA NTA Workflow

- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- Predicted Formulas
- Candidate Structures
- Prioritized Structures
- **Final Proposed Structures**
- Predicted Concentrations

Library TCP Fragmentation Spectra (20eV)



Observed Fragmentation Spectra (20eV)

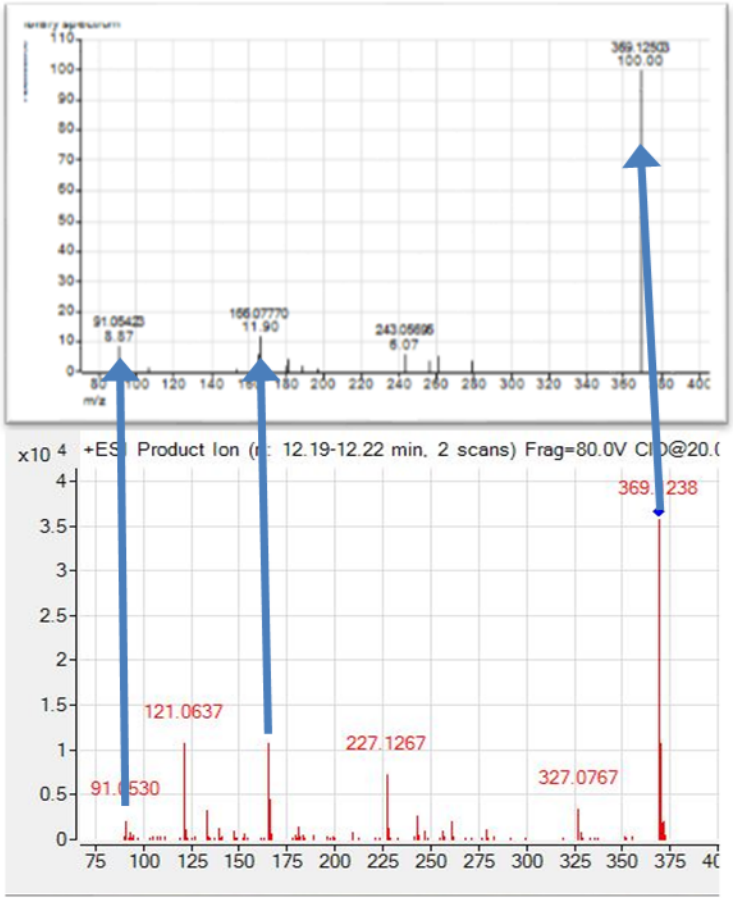
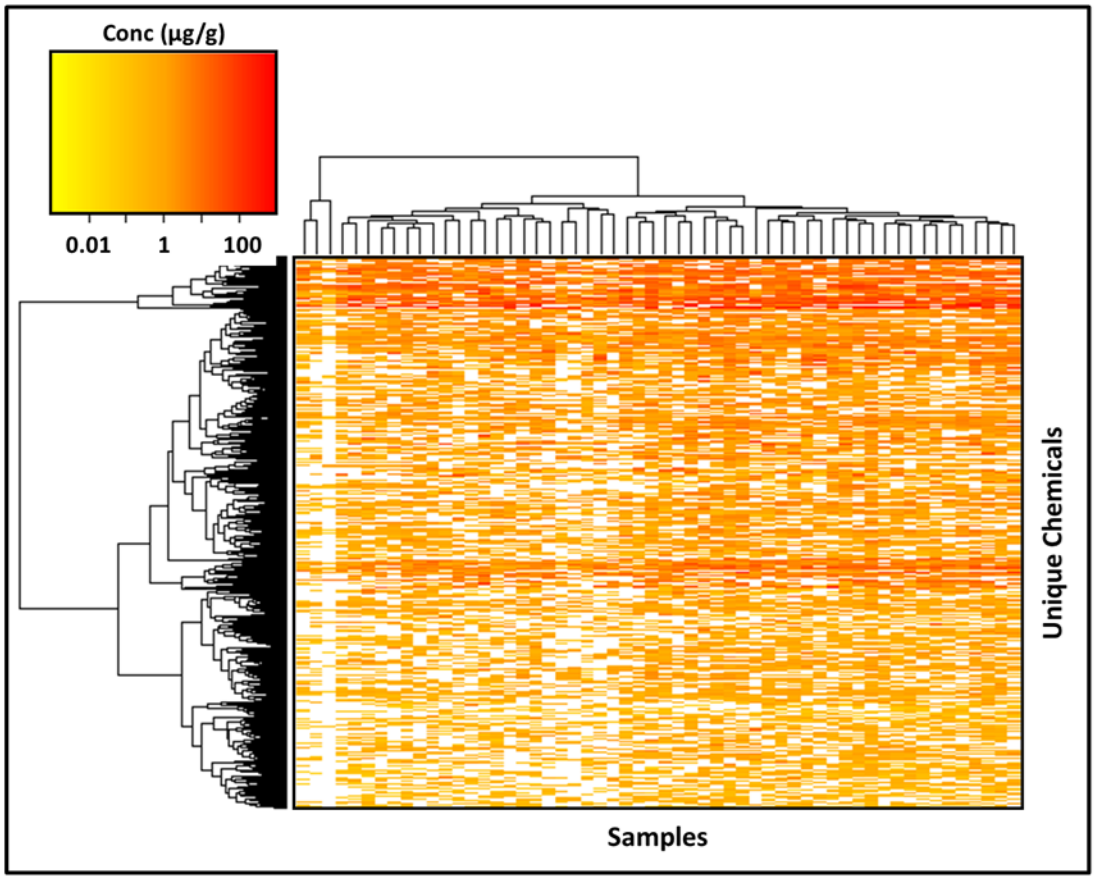


Figure from Jarod Grossman

EPA NTA Workflow

- Raw Samples
- Extracted Samples
- Raw Features
- Processed Features
- Predicted Formulas
- Candidate Structures
- Prioritized Structures
- Final Proposed Structures
- **Predicted Concentrations**



Applications of NTA Workflow

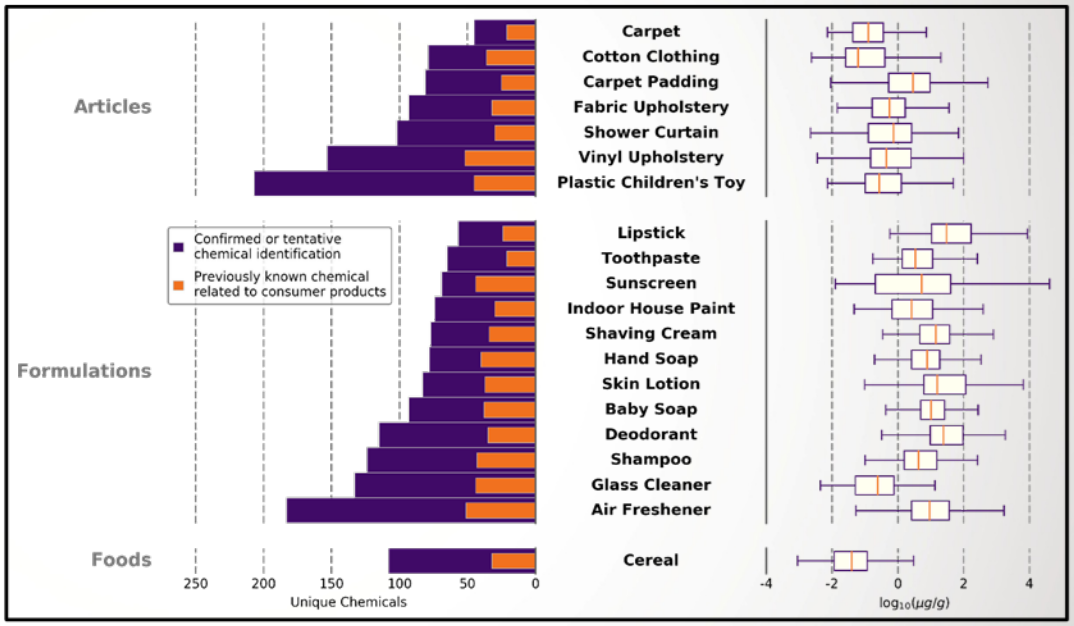
- **Exposure surveillance**
  - What chemicals are in food, products, dust, blood, etc.?
- **Chemical prioritization**
  - What are the most important chemicals & mixtures?
- **Exposure forensics**
  - What are chemical signatures of exposure sources?
- **Effect-directed analysis**
  - What are the biologically active chemicals in complex mixtures?
- **Biomarker discovery**
  - What chemicals are predictive of bioactivity/health impairment?

Product Analysis Pilot Study

20 product categories, 5 products per category

(12 formulation cat. + 7 article cat. + 1 food cat.)

Product Category	Number of Chemicals Identified: Ingredient List	Number of Chemicals Identified: Tentative NTA Hits
air freshener	4	183
baby soap	9	94
deodorant	6	115
glass cleaner	4	133
hand soap	10	79
lipstick	14	54
shampoo	10	125
shaving cream	9	78
skin lotion	10	80
sunscreen	7	69
toothpaste	6	66

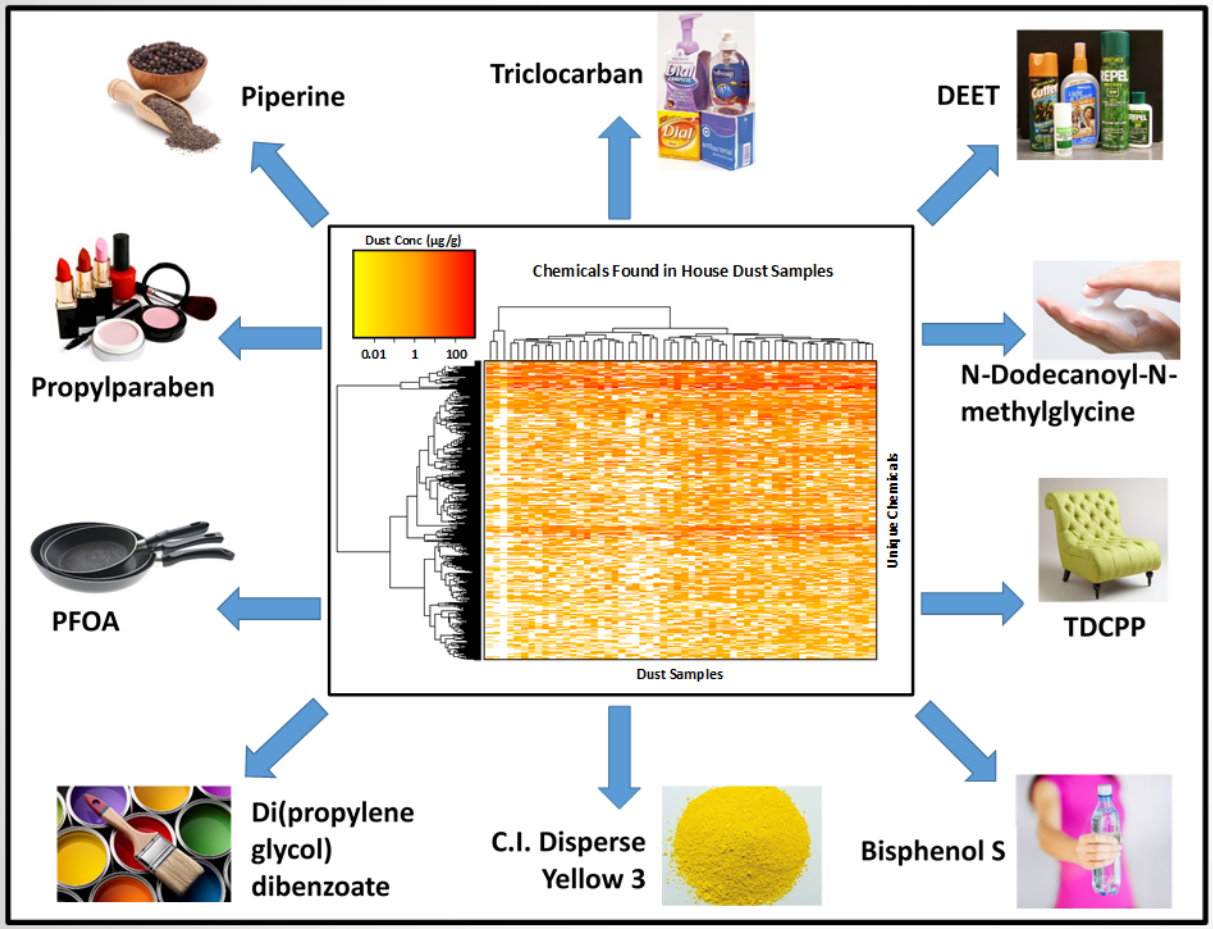


19% of chemicals identified by NTA are on consumer product chemical lists

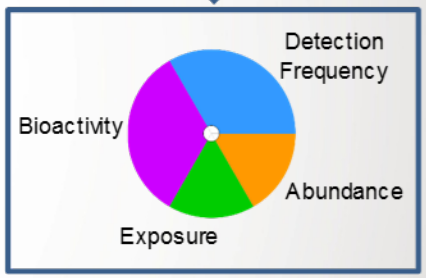
material from Katherine Phillips, submitted

House Dust Pilot Study

Dust samples from 56 homes (American Health Homes Survey)



Thousands of compounds prioritized using ToxPi



45% of confirmed compounds never before measured in dust

Identifying Tracers of Indirect Dust Ingestion

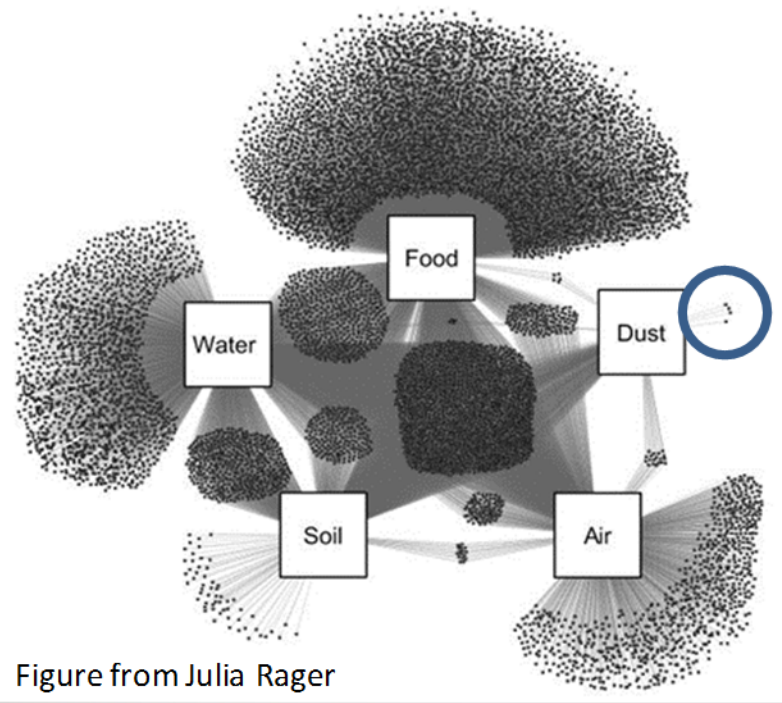
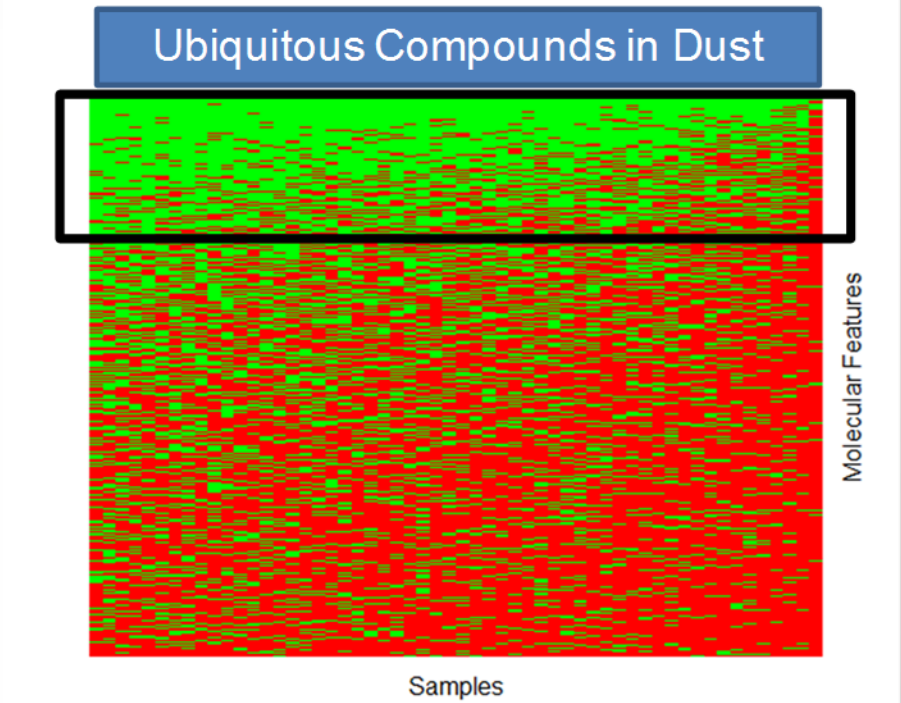


Figure from Julia Rager



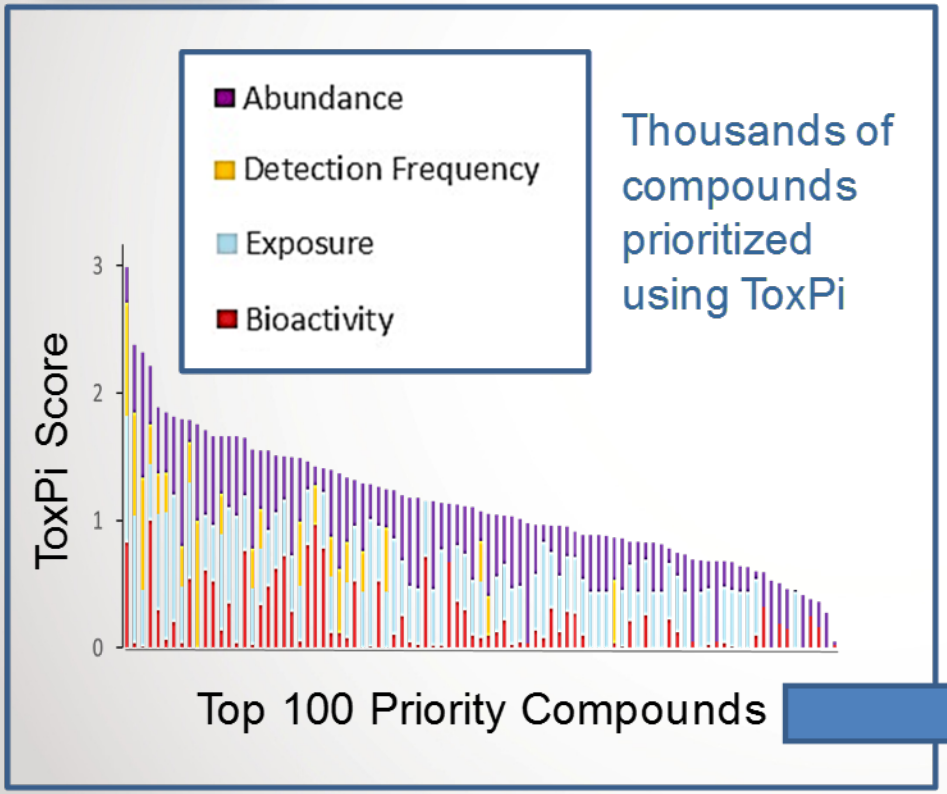
<u>Dust Ingestion Tracers Must Be</u>:

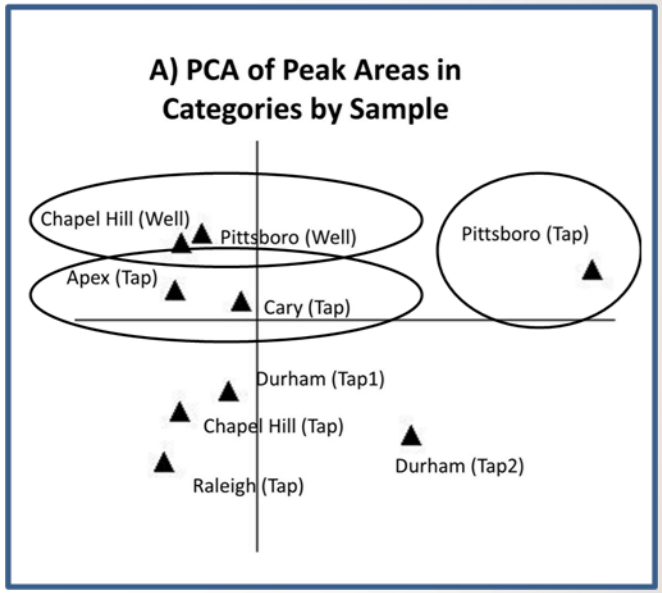
1) Ubiquitous in dust
2) Unique to dust
3) Excreted in an accessible medium
4) Measurable as a biomarker

NTA is suitable for evaluating candidates based on each of these criteria!

Drinking Water Pilot Study

NC Triangle Area:
- 9 samples representing 4 municipalities & 2 wells





-15 out of top 100 confirmed
- 8 out of top 20 confirmed
- Several not on monitoring lists
- Reflects consumer uses+ industrial processes

material from Seth Newton, submitted

Lingering Science Questions for NTA

- How variable are tools and results from lab to lab?
- Are some methods/tools better than others?
- Does sample complexity affect performance?

- What chemical space does a given method cover?
- Can we comprehensively characterize any sample?
- How does chemical content vary across media?

- To what extent do MS spectra aid in identification?
- How do we make MS spectra widely available?
- Can we crowdsource exposure data for ToxCast chemicals?

Lingering Science Questions for NTA

- How variable are tools and results from lab to lab?
- Are some methods/tools better than others?
- **EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)**
- To what extent do MS spectra aid in identification?
- How do we make MS spectra widely available?
- Can we crowdsource exposure data for ToxCast chems?

Samples, Participants and Products

ToxCast Chemicals

Reference House Dust

Reference Human Serum

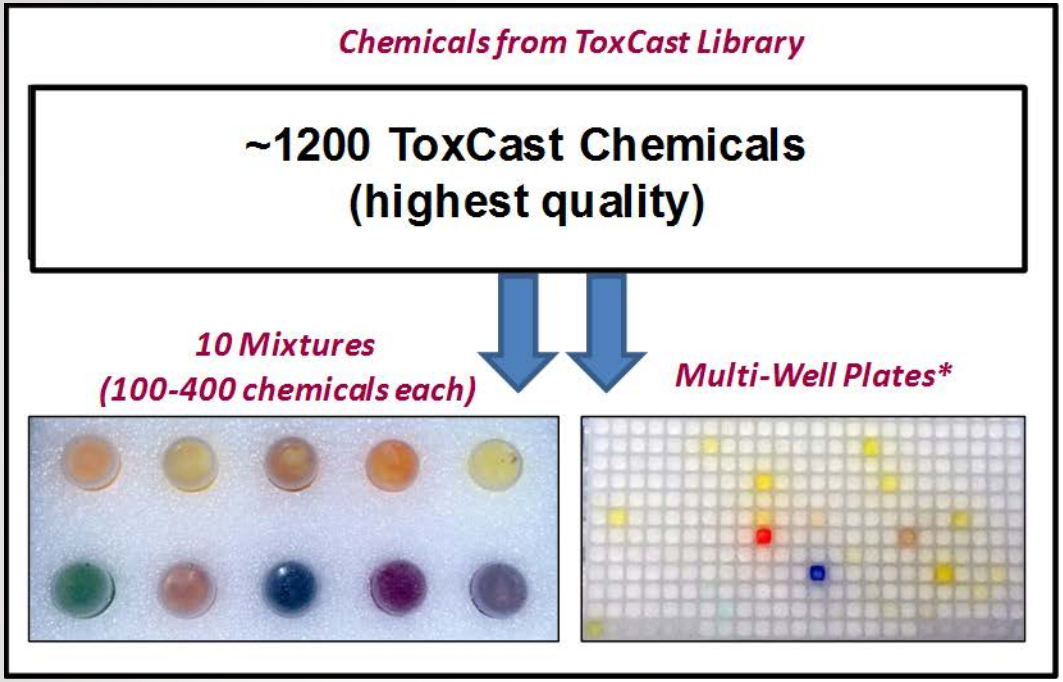
Reference Silicone Wristbands

Method Evaluation

Reference Libraries

Model Training Sets

Part 1



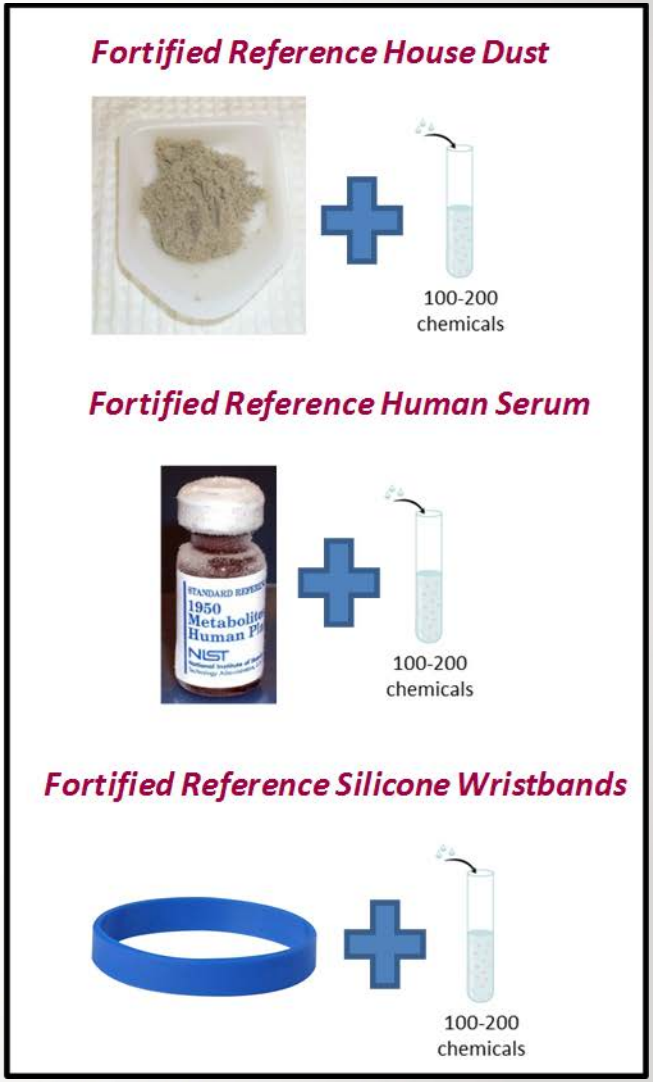
~20 Collaborators & 5 Contractors*:

1st: Blinded analysis

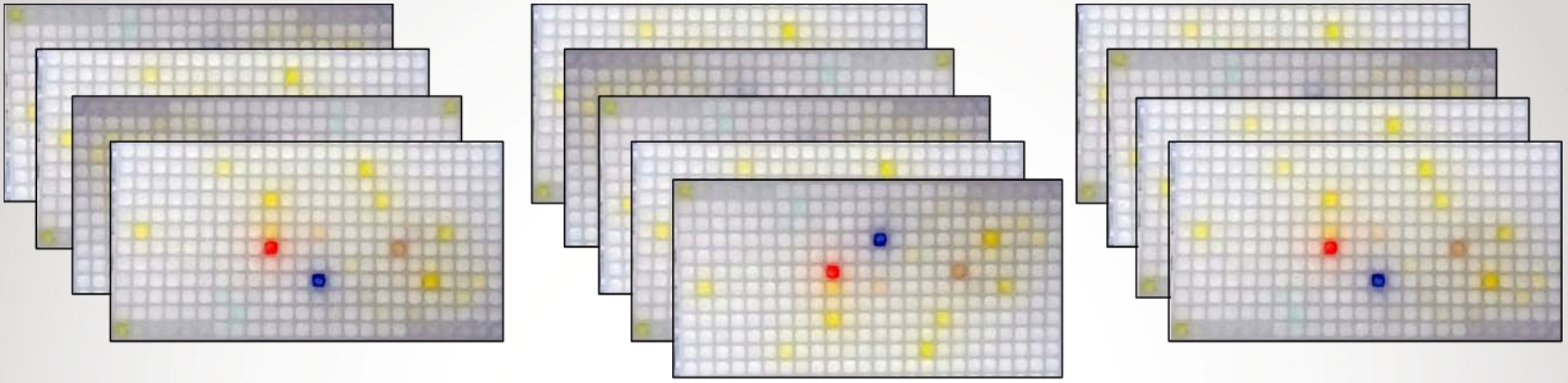
2nd: Unveiling of chemicals

3rd: Unblinded evaluation

Part 2



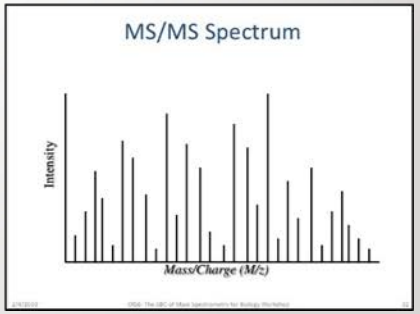
Part 3



~4600 ToxCast substances

Instrument/software vendors & select labs

Reference libraries for the public



Take-home Points

- NTA methods must become mainstream
- Currently no standardized methods/techniques
- EPA is working to establish benchmarks
  - Using ToxCast chemicals to test NTA methods
  - Building/sharing chemistry databases
  - Developing tools for rapid & accurate characterization
  - Integrating data with international NTA community
  - Coordinating international collaborative trial
  - Discovery, surveillance, forensics & prioritization

Acknowledgements

EPA NERL
Derya Biryol*
Kathie Dionisio
Jarod Grossman*
Kristin Isaacs
Sarah Laughlin*
Aurelie Marcotte*
James McCord*
Rebecca McMahen*
Seth Newton
Katherine Phillips
Julia Rager*
Mark Strynar
Elin Ulrich

EPA NCCT
Chris Grulke
Richard Judson
Kamel Mansouri*
Andrew McEachran*
Ann Richard
John Wambaugh
Antony Williams

EPA OSA
Marie Russell*

* = ORISE/ORAU/ASPPH

Web Art Links

- **Iceberg:** https://vimby.com/wp-content/uploads/2015/01/iceberg.jpg
- **LC-ESI-TOF:** https://masspec.scripps.edu/instruments/images/Agilent-ESI-TOF.jpg
- **GC Q-TOF:** http://www.agilent.com/cs/publishingimages/7200B_front_730x730.png
- **Orbitrap:** http://planetorbitrap.com/data/fe/image/QExactive.jpg
- **Black pepper 1:** https://sc02.alicdn.com/kf/UTB8mBZpXwnJXKJkSaelq6xUzXXae/Quality-Ground-Black-Pepper-for-Industrial-goods.jpg
- **Tire crumb**: http://assets.change.org/photos/5/hq/kx/zSHQkxioHipfgDd-1600x900-noPad.jpg?1426174444
- **Brita filter:** https://www.brita.com/wp-content/uploads/faucet-hero1.png
- **Minnow:** http://fishing.boyslife.org/files/2013/03/minnow-feature-620x465.jpg
- **Burning chair:** http://www.motherjones.com/wp-content/uploads/burningchair.jpg?w=990
- **Consumer products:** http://coreyko.fatcow.com/OC/wp-content/uploads/2013/10/CONSUMER-PRODUCTS.jpg
- **Blood:** https://ichef.bbci.co.uk/news/660/media/images/79490000/jpg/_79490509_135018352.jpg
- **Soil:** https://bonnieplants.com/wp-content/uploads/2011/12/soil-in-hands.jpg
- **Urine:** https://cdn.shopify.com/s/files/1/0996/0350/products/urine-specimen-cup__44522_0924aa8b-864a-47df-936f-58d9cebdbf21.jpeg?v=1499751369
- **Wristband:** https://www.wristbandbuddy.com/blog/wp-content/uploads/2016/03/Cheap-Silicone-Bracelets.jpg
- **Water effluent:** http://nts-industrie.com/wp-content/uploads/sites/2/2015/09/photo-traitement-de-leaux4-200x300.jpg
- **Dust:** https://www.ecocleanmadison.com/wp-content/uploads/2016/11/dust.jpg
- **Pilot:** https://upload.wikimedia.org/wikipedia/commons/thumb/9/9e/Defense.gov_News_Photo_060803-F-2907C-107.jpg/220px-Defense.gov_News_Photo_060803-F-2907C-107.jpg
- **Personal care products:** https://nourishingtransformations.com/wp-content/uploads/personal-care-products.jpg

Web Art Links

- **Black Pepper 2:** https://sc01.alicdn.com/kf/UT8E_BLX_xXXXagOFbXi/Organic-Black-Pepper-Powder-in-bulk-.jpg
- **Soaps:** http://www.bbcleaningservice.com/wp-content/uploads/2011/10/AntibacterialSoaps.jpg?x98918
- **Insect repellant:** http://labs.russell.wisc.edu/mosquitosite/files/2012/05/Deet-small.jpg
- **Cosmetics:** https://il8.picdn.net/shutterstock/videos/4062139/thumb/1.jpg
- **Foam soap:** https://pureessentialsupply.com/wp-content/themes/pureessentialsupply/shopp/images/Foaming%20soap.jpg
- **Non-stick pans:** https://cdn1.healthambition.com/wp-content/uploads/2017/02/28094108/Teflon-Dangers3.jpg
- **Chair:** https://s-media-cache-ak0.pinimg.com/564x/dd/c0/1d/ddc01d0078185b800e80df51a8afcb73.jpg
- **Yellow dye:** http://sunnychemi.com/wp-content/uploads/2016/06/product-3-550x410.jpg
- **Water bottle:** https://img.washingtonpost.com/wp-apps/imrs.php?src=https://img.washingtonpost.com/news/to-your-health/wp-content/uploads/sites/26/2015/01/bigstock-sport-and-diet-concept-sport-49821368.jpg&w=480
- **SRM plasma:** https://www.nist.gov/sites/default/files/styles/220_x_220_limit/public/images/mml/csd/srm-1950_1.jpg?itok=E8aCgLDJ
- **MSMS spectrum**: https://image.slidesharecdn.com/ms-100204133503-phpapp02/95/mass-spectrometry-protein-identification-strategies-32-728.jpg?cb=1265290569