http://orcid.org/0000-0003-1423-330X

EPA
United States
Environmental Protection
Agency

Using the US EPA's CompTox Chemistry Dashboard to advance non-targeted analysis and exposure research

Andrew D. McEachran

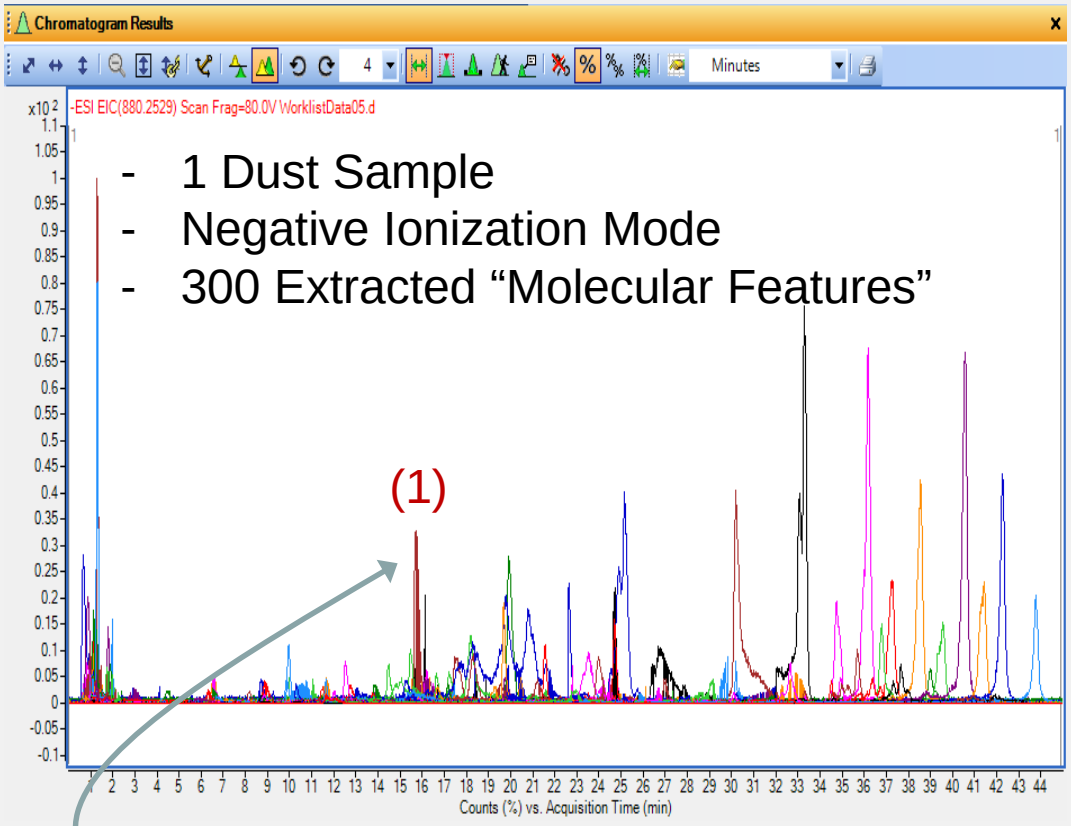
Jarod Grossman, Seth Newton, Kristin Isaacs, Katherine Phillips, Nancy Baker, Christopher Grulke, Jon R. Sobus, and Antony J. Williams

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

Office of Research and Development
National Center for Computational Toxicology, RTP, NC

October 16, 2017

General Goals of NTA



- 1 Dust Sample
- Negative Ionization Mode
- 300 Extracted "Molecular Features"

(1)

1) Prioritize "molecular features"

2) Correctly assign formulas

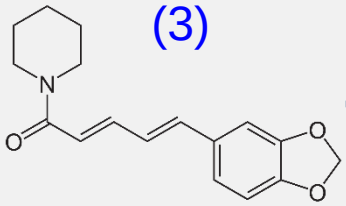
3) Correctly assign structures

4) Predict chemical concentrations

5) Determine chemical sources

(2) $C_{17}H_{19}NO_3$

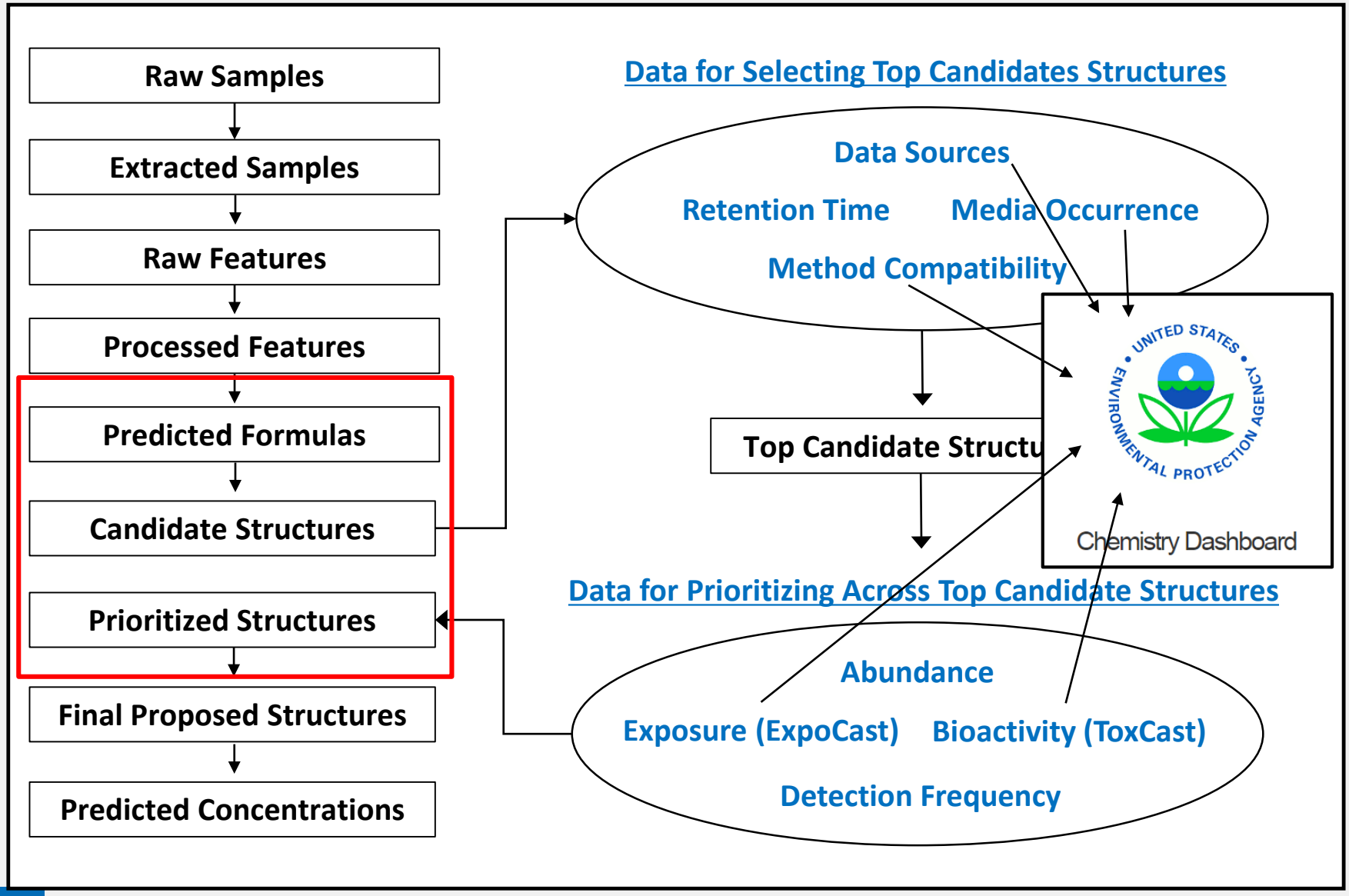
(3)



(4) 12 µg/g

(5)

EPA NTA Workflow



From J. Sobus

EPA United States Environmental Protection Agency

Chemical Properties | Env. Fate/Transport | Toxicity Values (Beta) | ADME (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Synonyms | Literature | External Links | Comments

Batch Search

Please enter one identifier per line

Select Input Type(s)

- [ ] Chemical Name
- [ ] CAS-RN
- [ ] InChIKey
- [ ] DSSTox Substance ID
- [ ] Exact Molecular Formula

Enter Identifiers to Search

Display All Chemicals | Download Chemical Data

EPA United States Environmental Protection

Batch Search

Please enter one identifier per line

Select Input Type(s)

- [ ] Chemical Name
- [ ] CASRN
- [ ] InChIKey
- [ ] DSSTox Substance ID
- [x] Exact Molecular Formula

Include top 10 hits in download

Enter Identifiers to Search

C13H18O2
C8H10N4O2
C6H6O

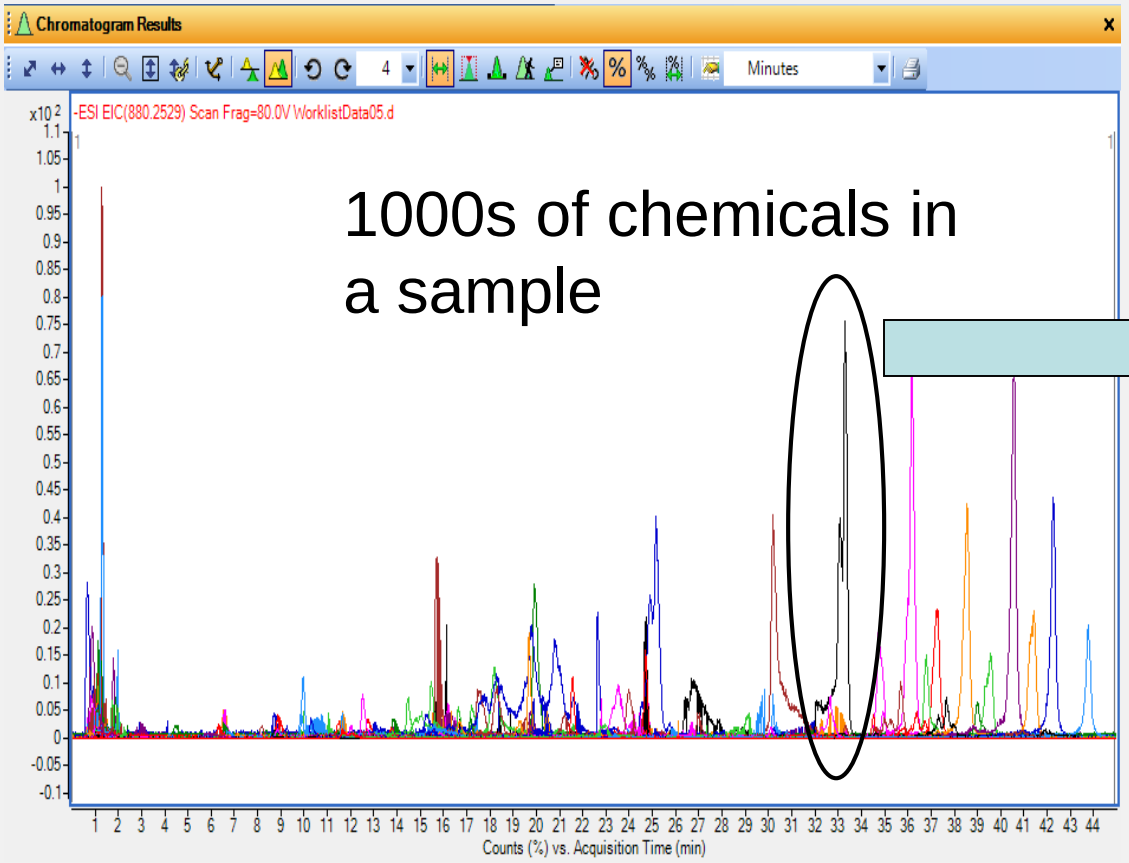
Display All Chemicals | Download Chemical Data

Metadata

- [ ] Curation Level Details
- [x] Data Sources
- [x] Assay Hit Count
- [ ] Include links to ACToR reports - SLOW! (BETA)
- [x] NHANES/Predicted Exposure
- [ ] Include ToxVal Data Availability

Presence In List

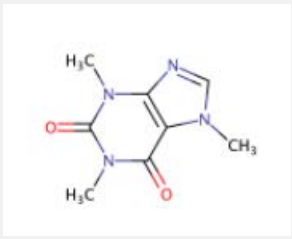
- [ ] Algal Toxins
- [ ] All chemicals in ChemTrack
- [ ] ATSDR Toxic Substances Portal Chemical List
- [ ] Bisphenol Compounds
- [ ] California Office of Environmental Health Hazard Assessment
- [ ] Chemicals with interesting names
- [ ] CMAP
- [x] DNT Screening Library
- [x] Drinking Water Suspects, KWR Water, Netherlands
- [ ] EDSP Universe
- [ ] EPA Hydrofracturing Fluids
- [ ] EPA Integrated Risk Information System (IRIS)
- [ ] EU Cosmetic Ingredients Inventory (Combined 2000/2006)
- [ ] EU Toxrisk Dataset
- [ ] French Monitoring List
- [ ] GENRA
- [ ] Grace Patlewicz's Chemicals of Interest
- [ ] HERO: Health and Environmental Research Online
- [ ] ITN ANTIBIOTIC LIST
- [ ] Joshua Harrill's HTTR_MCF7 List
- [ ] KEMI List of Substances on the Market
- [x] List of Swiss Pesticides and Transformation Products

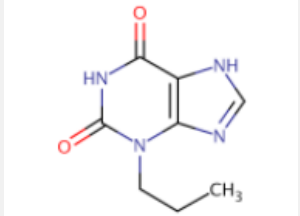


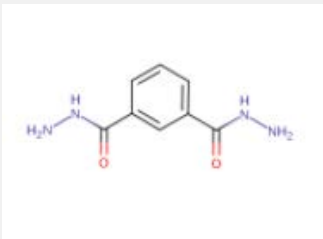
1000s of chemicals in a sample

1 formula

100s of structures



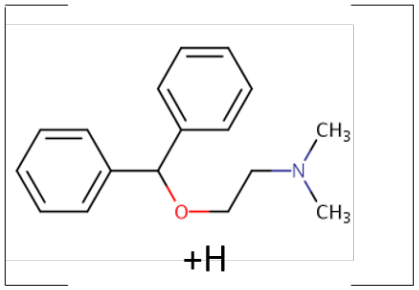




....?

MS-Ready structures improve database searching

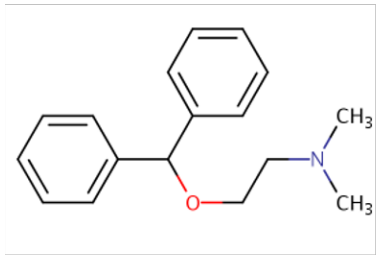
A) Molecular Ion



m/z≈ 256.170

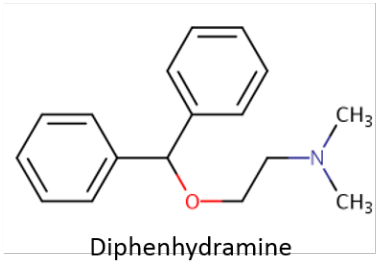
+

B) MS-Ready Form

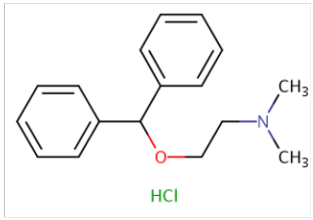


monoisotopic mass= 255.162
$C_{17}H_{21}NO$
DTXCID802949

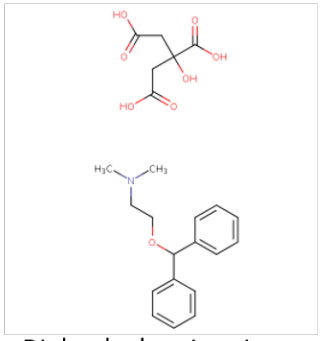
C) Mappings from MS-Ready



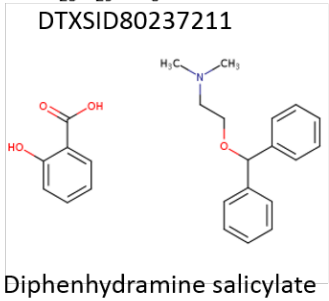
Diphenhydramine
$C_{17}H_{21}NO$
DTXSID4022949



Diphenhydramine
hydrochloride
$C_{17}H_{22}ClNO$
DTXSID4020537



Diphenhydramine citrate
$C_{23}H_{29}NO_8$
DTXSID80237211



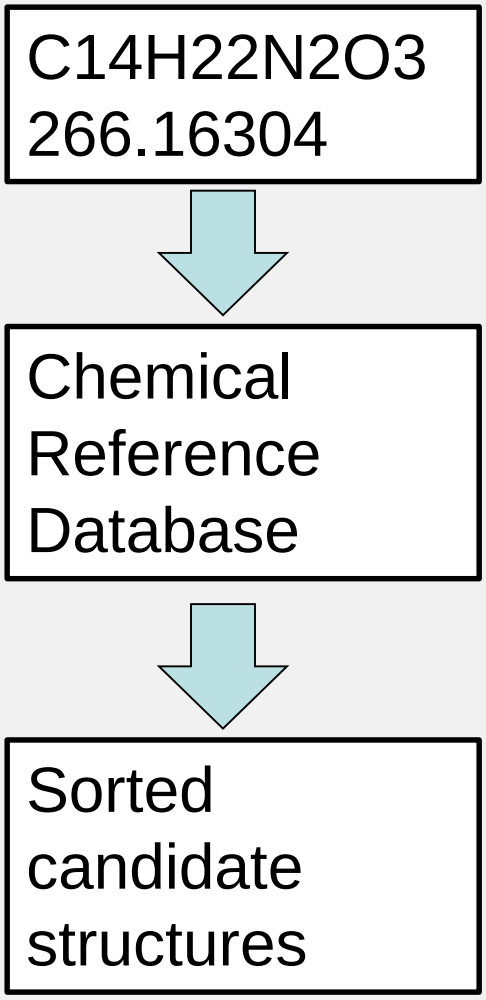
Diphenhydramine salicylate
$C_{24}H_{27}NO_4$
DTXSID10225883

McEachran, *et al*, in prep

EPA United States Environmental Protection Agency

Identification using data source ranking

- Mass and/or formula unknown to a researcher, contained within a reference database
- Most likely candidate chemicals have the most references/data sources



Data Source ranking brings most likely candidate structures to the top

- On same 162 chemicals, Dashboard outperforms ChemSpider

Anal Bioanal Chem
DOI 10.1007/s00216-016-0139-z

RAPID COMMUNICATION

Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran[1] · Jon R. Sobus[2] · Antony J. Williams[3]

	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2[a]	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

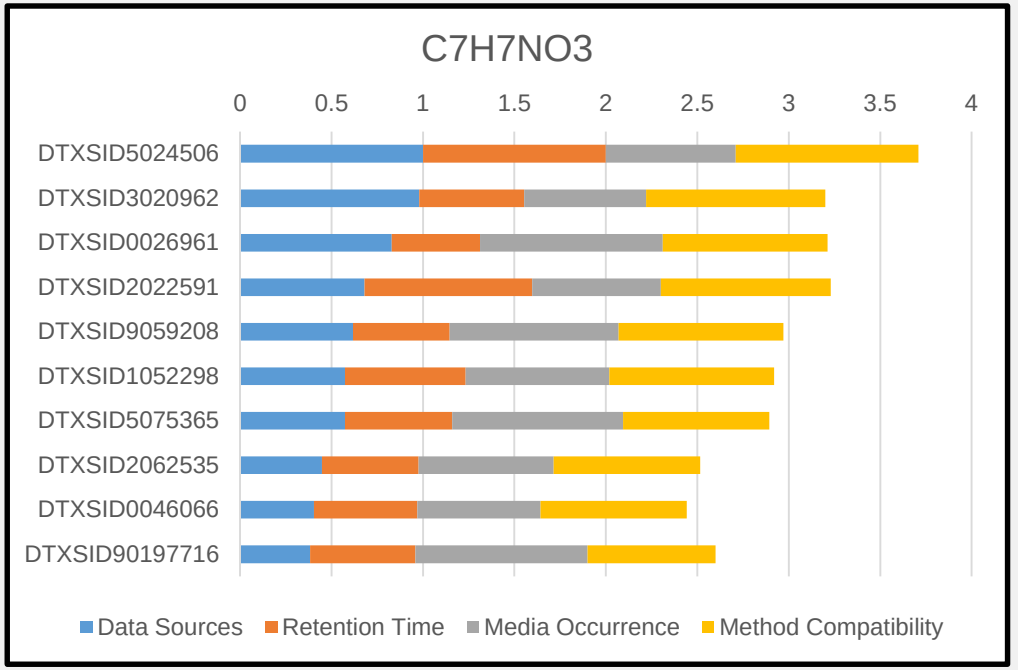
[a] Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5

DOI: 10.1007/s00216-016-0139-z

Incorporating additional data streams reduces uncertainty in identification

- US EPA CompTox Chemistry Dashboard Data Sources
- PubChem Data Source Count
- PubMed Reference Count
- Retention Time Prediction
- Predicted Environmental Media Occurrence
- OPERA PhysChem Properties
- Method Compatability
- ……

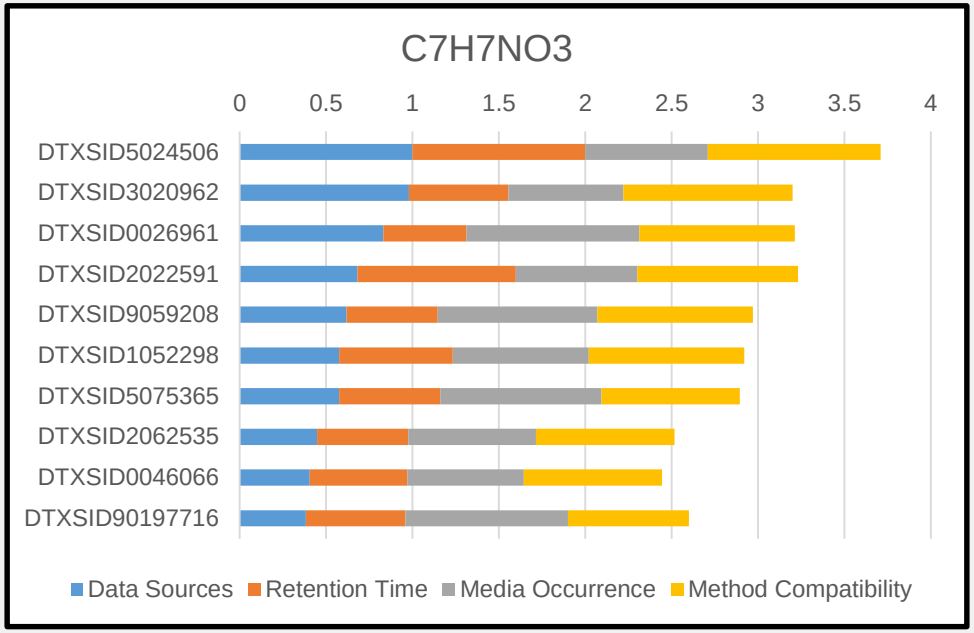
$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \cdots$$



Incorporating additional data streams increases the number of identifications

- Test Sets:
- ~700 chemicals occurring in surface water (Schymanski et al)
  - 3% increase over data sources alone/ 94% in top 5
- ~200 chemicals in dust (NIST SRM 2585)
  - 5% increase/ 99% in top 5

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \cdots$$

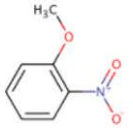
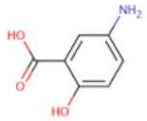
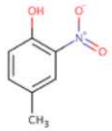
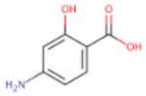


Incorporating scoring scheme into results

Search Results

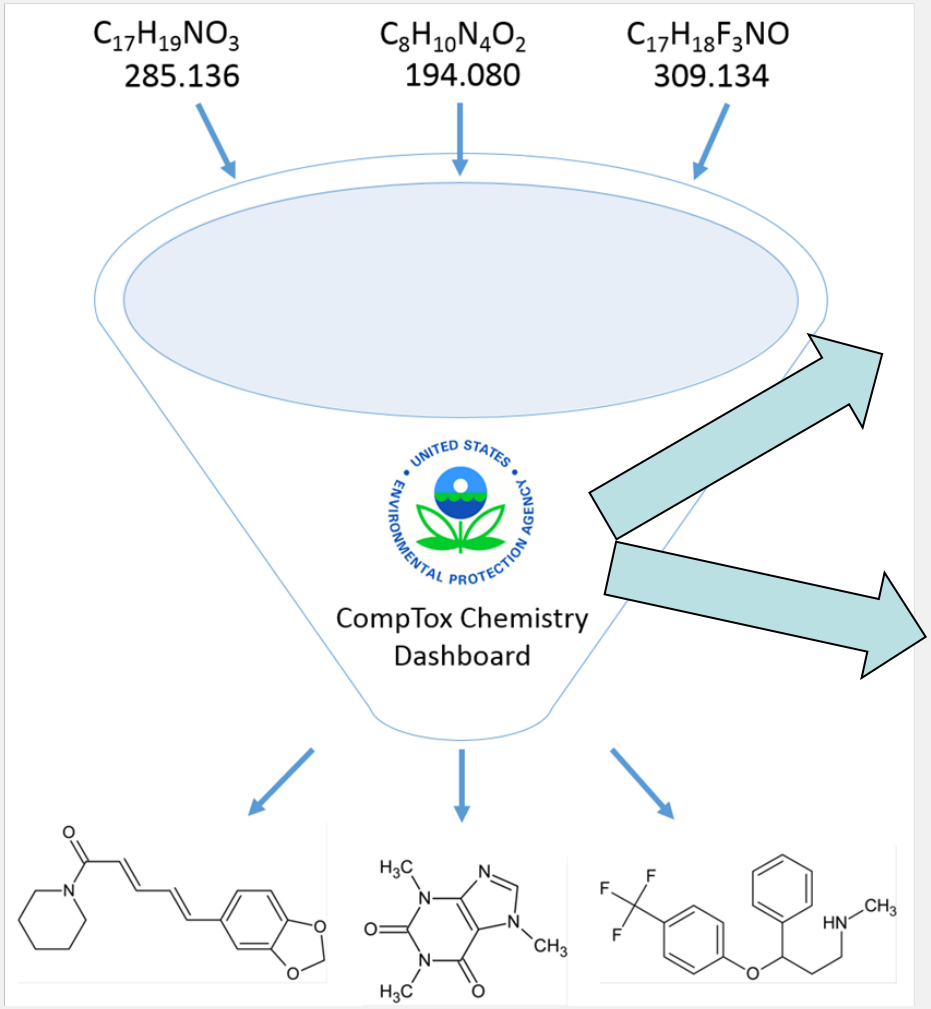
Searched by molecular formula: Found 188 results.

Download as: TSV | Excel | SDF

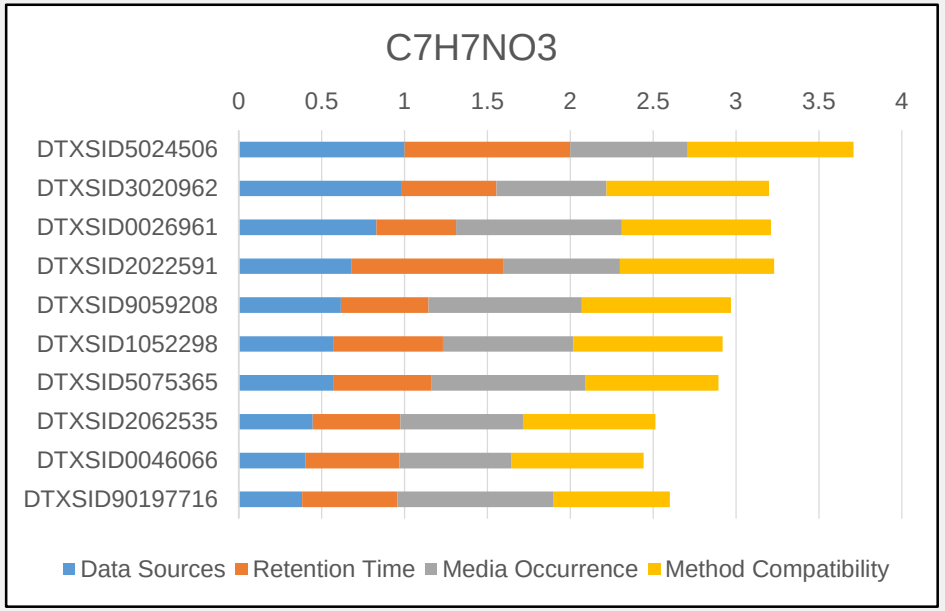
ID	Structure	Preferred Name	CAS-RN	QC Level	CPCat Count	Number of S...	PubChem D...	Monoisotopic Mass	Identification Score
									Data Sources · PubMed Ct · Media Occurr · CPDat Ct
DTXSID3020962 ToxCast™		1-Methoxy-2-nitrobenzene	91-23-6	Level 2: Expert curate...	0	58	111	153.042593	3.6
DTXSID5024506 ToxCast™		5-Aminosalicylic acid	89-57-6	Level 2: Expert curate...	2	57	193	153.042593	3.3
DTXSID0026961 ToxCast™		4-Methyl-2-nitrophenol	119-33-5	Level 2: Expert curate...	0	46	93	153.042593	3.3
DTXSID2022591		Aminosalicylic acid	65-49-6	Level 1: Expert curate...	0	40	177	153.042593	2.5

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \cdots$$

United States Environmental Protection Agency

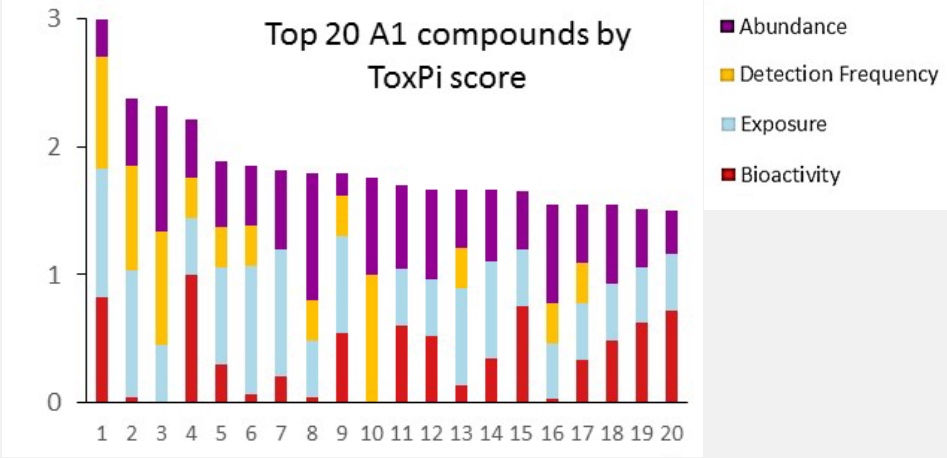


Identification



McEachran, *et al*

Prioritization



Newton, *et al* (2017)

Conclusions/future directions

- CompTox Chemistry Dashboard is a valuable open resource for exposure scientists
- NTA/SSA identification functionality has been developed and is currently available on the Dashboard
- Functionality demonstrates potential for chemical identification prioritization
- Continued workflow optimization is ongoing

Acknowledgements

EPA NCCT
Tony Williams
Nancy Baker
Chris Grulke
John Wambaugh
Kamel Mansouri*
Jeff Edwards
Ann Richard
Jennifer Smith

EPA NERL
Katherine Phillips
Kristin Isaacs
Kathie Dionisio
Jon Sobus
Mark Strynar
Elin Ulrich
Seth Newton
Jarod Grossman
Sarah Laughlin-Toth*
Aurelie Marcotte*
James McCord*

*ORISE Research Participant

EPA United States Environmental Protection Agency

Questions?

- mceachran.andrew@epa.gov
- http://orcid.org/0000-0003-1423-330X

- https://comptox.epa.gov/dashboard