

Non-targeted screening to improve substance identity for UVCBs

Antony J. Williams¹ and Emma L. Schymanski²

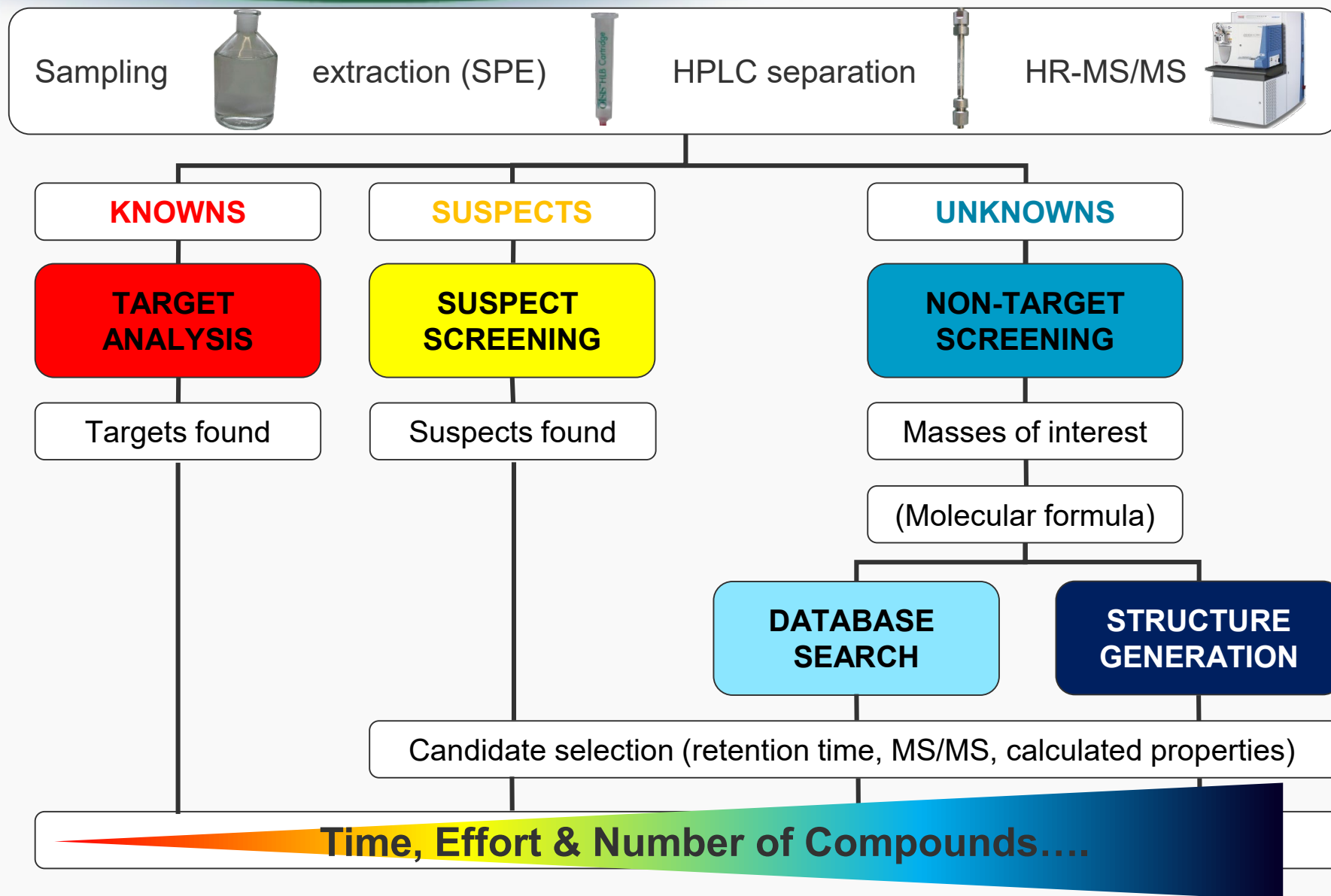
1) *National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*

2) *Luxembourg Center for Systems Biomedicine (LCSB), University of Luxembourg*

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

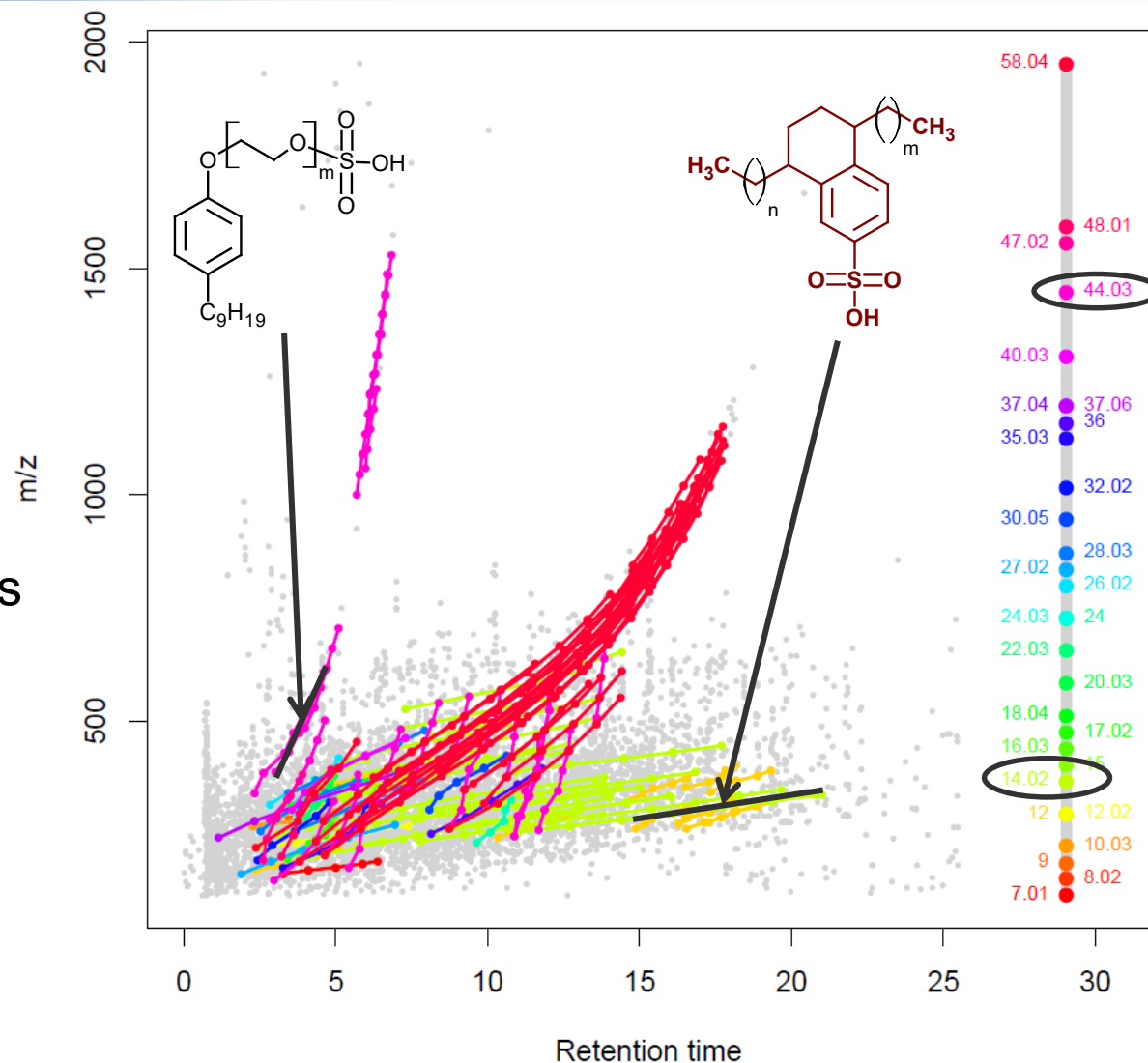
*December 2018
HESI Meeting on UVCB Chemicals*

What is Non-Target Analysis with Mass Spectrometry?

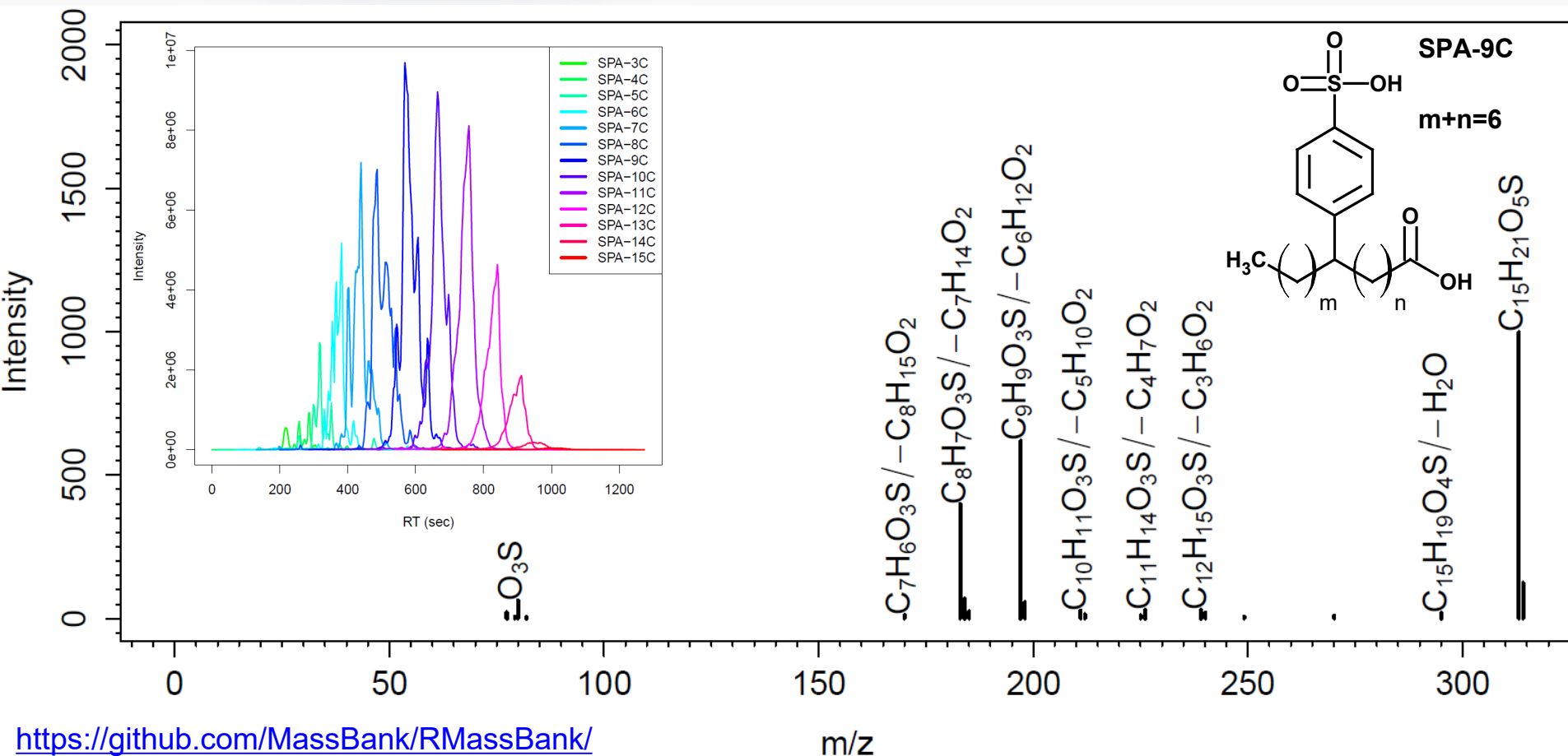


Detecting homologues series

- Search for discrete mass differences – homologs
- Provides evidence for series based relationships common in many UVCBs

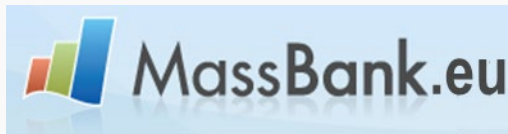


Supporting Evidence for Homologue Identification



<https://github.com/MassBank/RMassBank/>

Formulas: <http://sourceforge.net/projects/genform/>
Meringer *et al.* 2011, *MATCH* 65, 259-290
Data: Schymanski *et al.* 2014, *ES&T*, 48:
1811-1818. DOI: 10.1021/es4044374

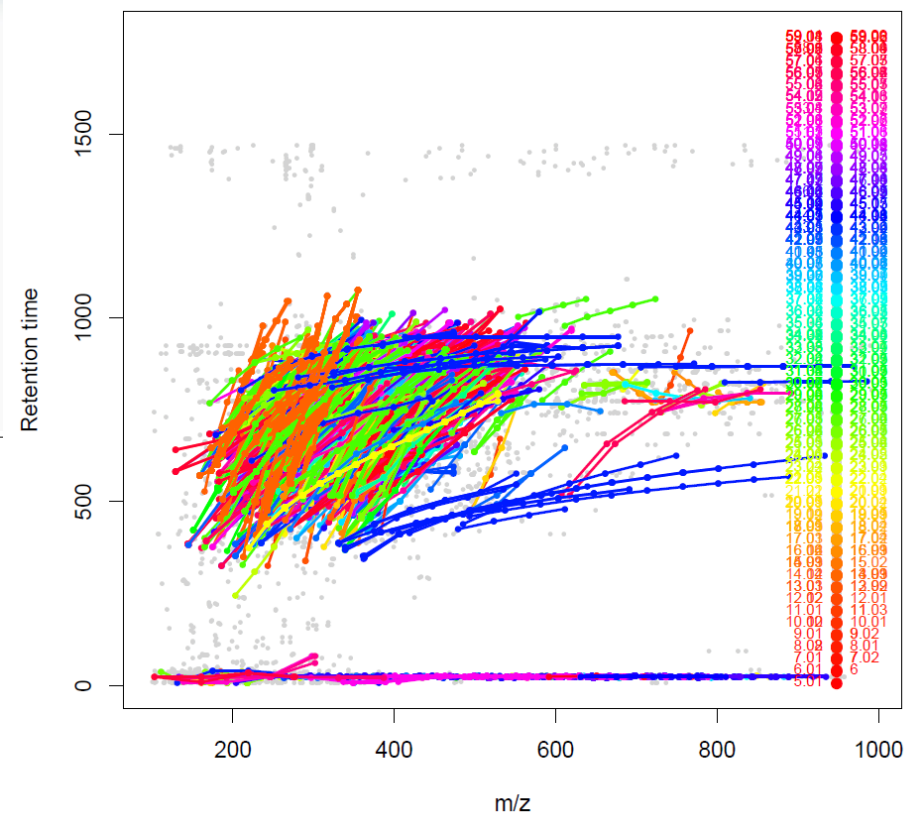
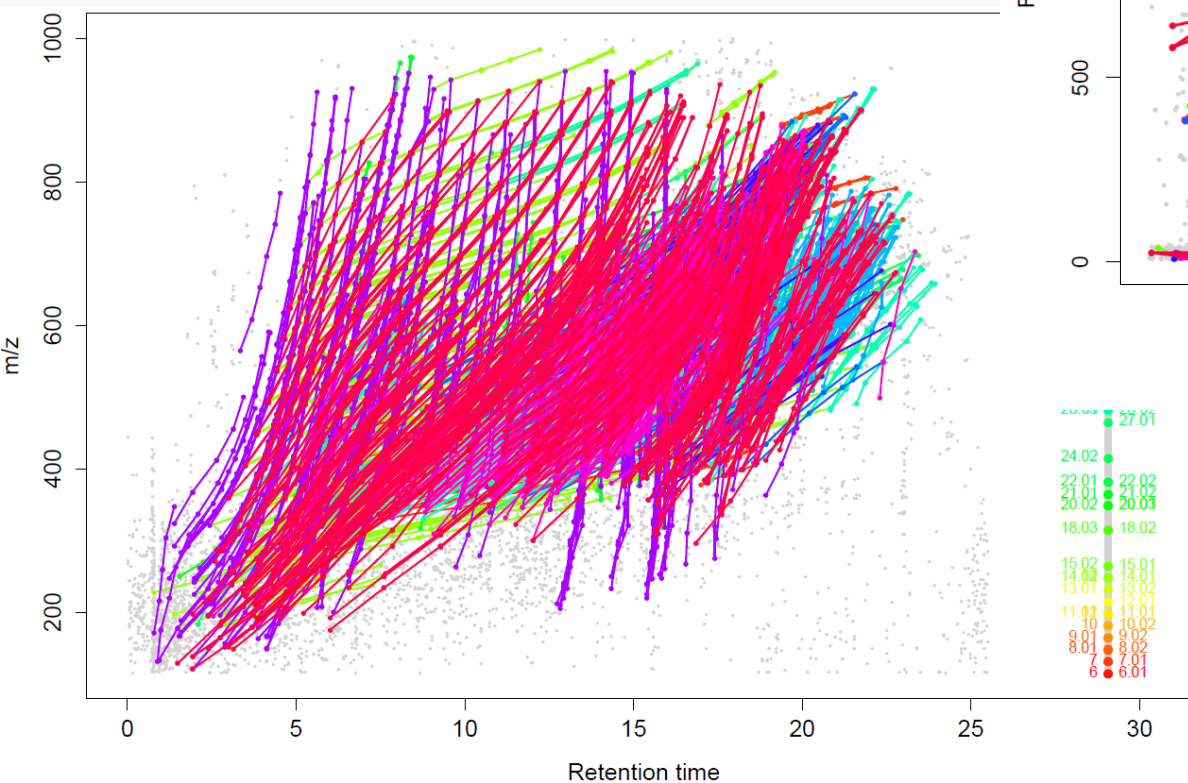


Literature: LIT00034,35
Sample: ETS00002
Standard: ETS00016,17,19,20

Stravs *et al.* (2013), *J. Mass Spectrom.* 48(1):89-99. DOI: 10.1002/jms.3131

Our challenge in non-target analysis

- Complex mixtures (UVCBs) are a **huge** and **very challenging** part of the unknowns in many environmental samples



Homologue screening plots from
Swiss Wastewater (Schymanski *et al*
2014, left) and Novi Sad (right)

solutions

EPA Approaches: The dashboard as the glue

Williams et al. / J Cheminform (2017) 9:61
DOI 10.1186/s13321-017-0247-6

Journal of Cheminformatics

DATABASE

Open Access

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry

Antony J. Williams^{1*}, Christopher M. Grulke¹, Jeff Edwards¹, Andrew D. McEachran², Kamel Mansouri^{1,2,4}, Nancy C. Baker³, Grace Delawie¹, Imran Shah¹, John F. Wambaugh¹, Richard S. Judson¹ and Ann M. Richard¹

Anal Bioanal Chem (2017) 409:1729–1735
DOI 10.1007/s00216-016-0139-z

RAPID COMMUNICATION

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

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³



Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-018-1435-6

RESEARCH PAPER

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

Elin M. Ulrich¹ · Jon R. Sobus¹ · Christopher M. Grulke² · Ann M. Richard² · Seth R. Newton¹ · Mark J. Strynar¹ · Kamel Mansouri^{3,4} · Antony J. Williams²

Received: 30 July 2018 / Revised: 14 September 2018 / Accepted: 17 October 2018
© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2018



Cite the Dashboard Publication [click here](#)

Journal of Exposure Science & Environmental Epidemiology
https://doi.org/10.1038/s41370-017-0012-y

REVIEW ARTICLE

Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA

Jon R. Sobus¹ · John F. Wambaugh² · Kristin K. Isaacs¹ · Antony J. Williams² · Andrew D. McEachran³ · Ann M. Richard² · Christopher M. Grulke² · Elin M. Ulrich¹ · Julia E. Rager^{3,4} · Mark J. Strynar¹ · Seth R. Newton¹



Talanta 182 (2018) 371–379

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

A comparison of three liquid chromatography (LC) retention time prediction models

Andrew D. McEachran^{a,b,*}, Kamel Mansouri^{a,b,1}, Seth R. Newton^c, Brandiese E.J. Beverly^{a,c,2}, Jon R. Sobus^c, Antony J. Williams^{b,*,**}



Chem

Se

Iden

Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

This paper is a compendium of information related to the broad class of chemical substances referred to as UVCBs for the Toxic Substances Control Act (TSCA) Chemical Substance Inventory. These chemical substances cannot be represented by unique structures and molecular formulas.

Public TSCA Inventory on Dashboard

18,696 Chemicals (12/9/2018)

TSCA Inventory, active non-confidential portion

☐ Substring search

List Details

Description: Section 8 (b) of the Toxic Substances Control Act (TSCA) requires EPA to compile, keep current and publish a list of each chemical substance that is manufactured or processed, including imports, in the United States for uses under TSCA. Information about what types of substances are on the TSCA inventory can be found [here](#). The Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, requires EPA to designate chemical substances on the TSCA Chemical Substance Inventory as either "active" or "inactive" in U.S. commerce. To accomplish this, EPA finalized a rule requiring industry reporting of chemicals manufactured (including imported) or processed in the U.S.. This reporting is used to identify which chemical substances on the TSCA Inventory are active in U.S. commerce and help inform the prioritization of chemicals for risk evaluation. The list contained in the dashboard includes the active TSCA inventory based on notifications through [Feb. 7th 2018](#) and substances reported from [Feb 8, 208 – March 30, 2018](#) that have been unambiguously mapped to DSSTox using CASRN and chemical names. The curation of the non-confidential portion of active TSCA inventory is an ongoing process involving trained chemists to validate the correctness of DSSTox structural and identifier data. The content of the list will change over time as both the non-confidential active TSCA inventory is updated and more substances are curated.

Number of Chemicals: 18696

18696 chemicals

Download / Send ▾

Show info:

DTXSID ×

PubChem ×

CASRN × ▾

Select all



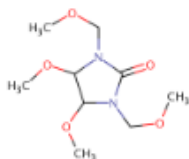
Sort by: DTXSID ▾



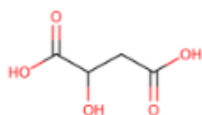
Filter by: Name or CASRN

Hide ▾

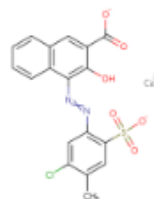
Many Chemicals are “Complex”



2-Imidazolidinone, 4,5-dimethoxy-1,3-bis(methoxymethyl)-
DTXSID: DTXSID0027569
PubChem: 24
CASRN: 4356-80-9



Malic acid
DTXSID: DTXSID0027640
PubChem: 273
CASRN: 6915-15-7



C.I. Pigment Red 48, calcium salt (1:1)
DTXSID: DTXSID0027642
PubChem: 0
CASRN: 7023-61-2

0 related chemical
structures with this
substance

Lard, oil
DTXSID: DTXSID0027690
PubChem: 0
CASRN: 8016-28-2

0 related chemical
structures with this
substance

Tall-oil pitch
DTXSID: DTXSID0027692
PubChem: 0
CASRN: 8016-81-7

0 related chemical
structures with this
substance

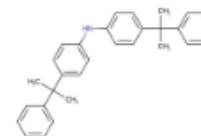
Palm kernel oil
DTXSID: DTXSID0027694
PubChem: 0
CASRN: 8023-79-8

0 related chemical
structures with this
substance

Tallow, hydrogenated
DTXSID: DTXSID0027696
PubChem: 0
CASRN: 8030-12-4

0 related chemical
structures with this
substance

Quaternary ammonium compounds, tri...
DTXSID: DTXSID0027698
PubChem: 0
CASRN: 8030-78-2



4-(2-Phenylpropan-2-yl)-N-[4-(2-phenyl...
DTXSID: DTXSID0027721
PubChem: 50
CASRN: 10081-67-1

1 related chemical
structure with this
substance

Isomethyltetrahydrophthalic anhydride
DTXSID: DTXSID0027729
PubChem: 0
CASRN: 11070-44-3

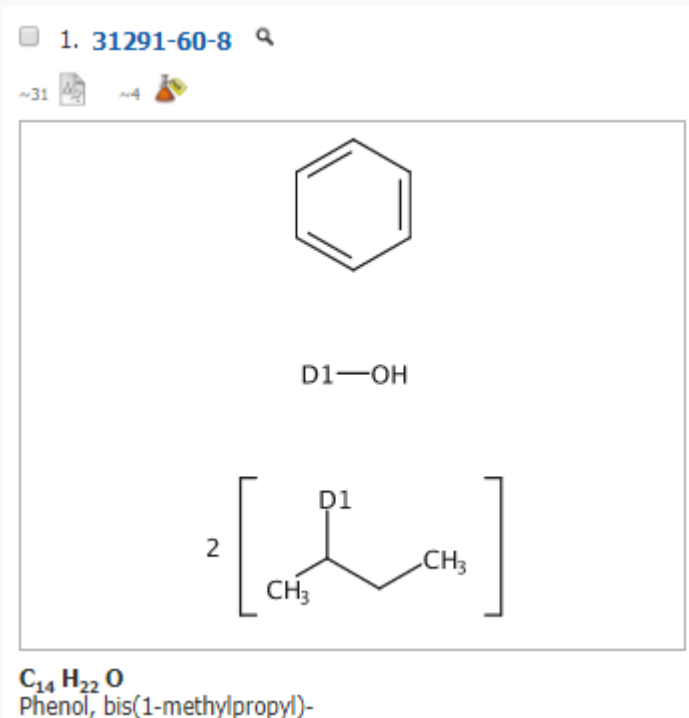
>7500 chemicals are UVCBs

0 related chemical structures with this substance	DTXSID00108838	2-Propenoic acid, 2-methyl-, polymer with N-(1,1-dimethyl-3-oxobutyl)-2-propenamide, ethyl 2-propenoate and methyl 2-methyl-2-propenoate, ammonium salt	173091-76-4
0 related chemical structures with this substance	DTXSID00108853	Tin, Bu 1-dodecanethiol 2-mercaptoethanol thioxo complexes	210920-40-4
0 related chemical structures with this substance	DTXSID00108858	1,4-Cyclohexanedicarboxylic acid, 1,4-dimethyl ester, trans-, polymer with 1,4-cyclohexanedimethanol	219586-57-1
0 related chemical structures with this substance	DTXSID00108878	Rosin, polymd., compd. with 2-(dimethylamino)ethanol	259191-53-2

Can these be represented?

Di-sec-butylphenol

CAS Representation



Dashboard Representation

Di-sec-butylphenol

31291-60-8 | DTXSID5049574

🔍 Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID5049574'.

Presence in Lists

Record Information

Quality Control Notes

Related Substances

Synonyms

Links

Bioassays

Exposure

Hazard

Comments

Chemical Properties

Literature

Download / Send

Sort by: Relationship



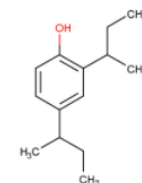
3 chemicals

Searched Chemical

2 related chemical
structures with this
substance

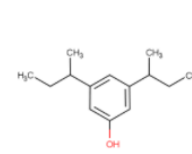
Di-sec-butylphenol
31291-60-8

Representative Isomer



2,4-Di-sec-butylphenol
1849-18-9

Representative Isomer



3,5-Bis(1-methylpropyl)phenol
14556-13-9

How to represent complexity?

Chemistry Dashboard

Submit Comment

Share ▾

Copy ▾

Aa ▾

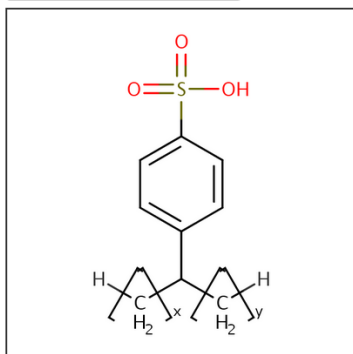
Aa

Aa ▴

Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

© Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID3020041'.



Intrinsic Properties

Molecular Formula: $(CH_2)_y(CH_2)xC_7H_8O_3S$

Find All Chemicals

Average Mass: Not Found

Monoisotopic Mass: Not Found

Structural Identifiers

Presence in Lists

Record Information

Quality Control Notes

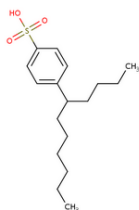
Related Substances

Chemical Properties

Analytical

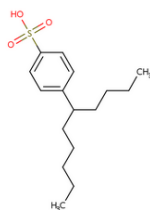
Comments

Markush Child



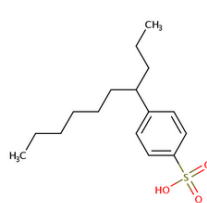
4-(undecan-5-yl)benzene-1-sulfonic acid
NOCAS_881097

Markush Child



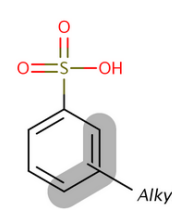
4-(decan-5-yl)benzene-1-sulfonic acid
NOCAS_881146

Markush Child



4-(decan-4-yl)benzenesulfonic acid
NOCAS_891333

Representative Component



(C10-C16) Alkylbenzenesulfonic acid
68584-22-5

Nested relationships

Download / Send

Sort by: Relationship



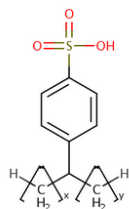
19 chemicals

Hide:

Select all



Searched Chemical



Alkylbenzenesulfonate, linear
42615-29-2

General Form

3 related chemical
structures with this
substance

Benzenesulfonic acid, C10-16-alkyl deri...
68584-25-8

General Form

2 related chemical
structures with this
substance

Benzenesulfonic acid, C10-13-alkyl deri...
68411-30-3

General Form

1 related chemical
structure with this
substance

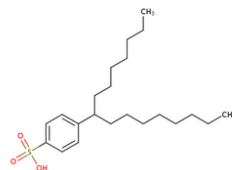
Benzenesulfonic acid, C10-16-alkyl deri...
68584-24-7

General Form

1 related chemical
structure with this
substance

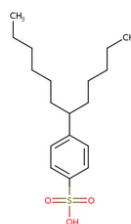
Benzenesulfonic acid, mono-C9-17-bra...
68649-00-3

Markush Child



4-(1-Heptylnonyl)benzenesulfonic acid
80233-94-9

Markush Child



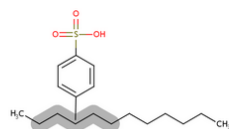
4-(Dodecan-6-yl)benzene-1-sulfonic acid
23003-92-1

Representative Component

4 related chemical
structures with this
substance

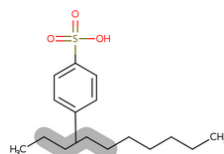
Benzenesulfonic acid, dodecyl-, branch...
90218-35-2

Representative Component



C12-linear alkyl benzene sulfonate
NOCAS_891641

Representative Component

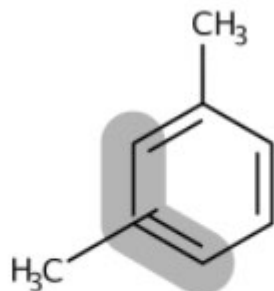


C10-linear alkylbenzenesulfonate
NOCAS_891689

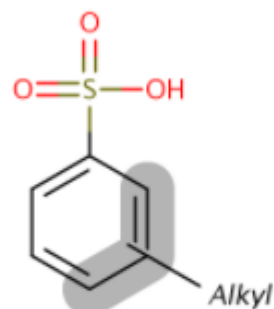
- 1) Structural relationships provides mass/formula connectivity
- 2) Individual components and series can be registered and related
- 3) Communication of UVCB structure details possible

“Markush Structures”

https://en.wikipedia.org/wiki/Markush_structure



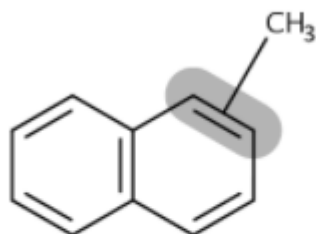
Xylenes
1330-20-7



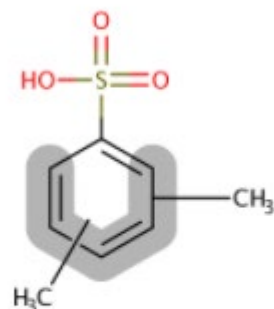
(C10-C16) Alkylbenzenesulfonic acid
68584-22-5



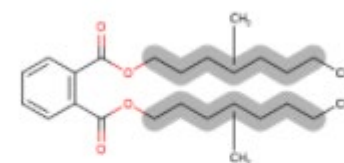
n-Nonylphenol
25154-52-3



Methylnaphthalene
1321-94-4



Sodium xylenesulfonate
1300-72-7



Diisononyl phthalate
28553-12-0

- Characterization of UVCB unknowns is possible using (various) MS approaches
- Many UVCBs have inherent mass/formula relationships as homologous series
- Need to link regulated mixtures without defined structures to observations in environment
- Cheminformatics approaches for complex UVCBs progressing with enumeration (Markush). Relationship mappings assist this
- Lots needed to progress open science informatics approaches but well underway...

- Contributions from many of our collaborators
 - European NORMAN Network
 - European MassBank
 - Open Source Cheminformatics – CDK, MetFrag
 - Eawag, IPB, UFZ, ETHZ and LCSB colleagues
 - EPA NERL colleagues
 - EPA NCCT colleagues