

# ToxCast: Chemical List(s), Chemical Space, and Chemotype Information

*Ann Richard  
U.S. EPA, National Center for Computational Toxicology  
Office of Research and Development*



CompTox Communities of Practice, February 27, 2014

# ToxCast Phase II Data Release

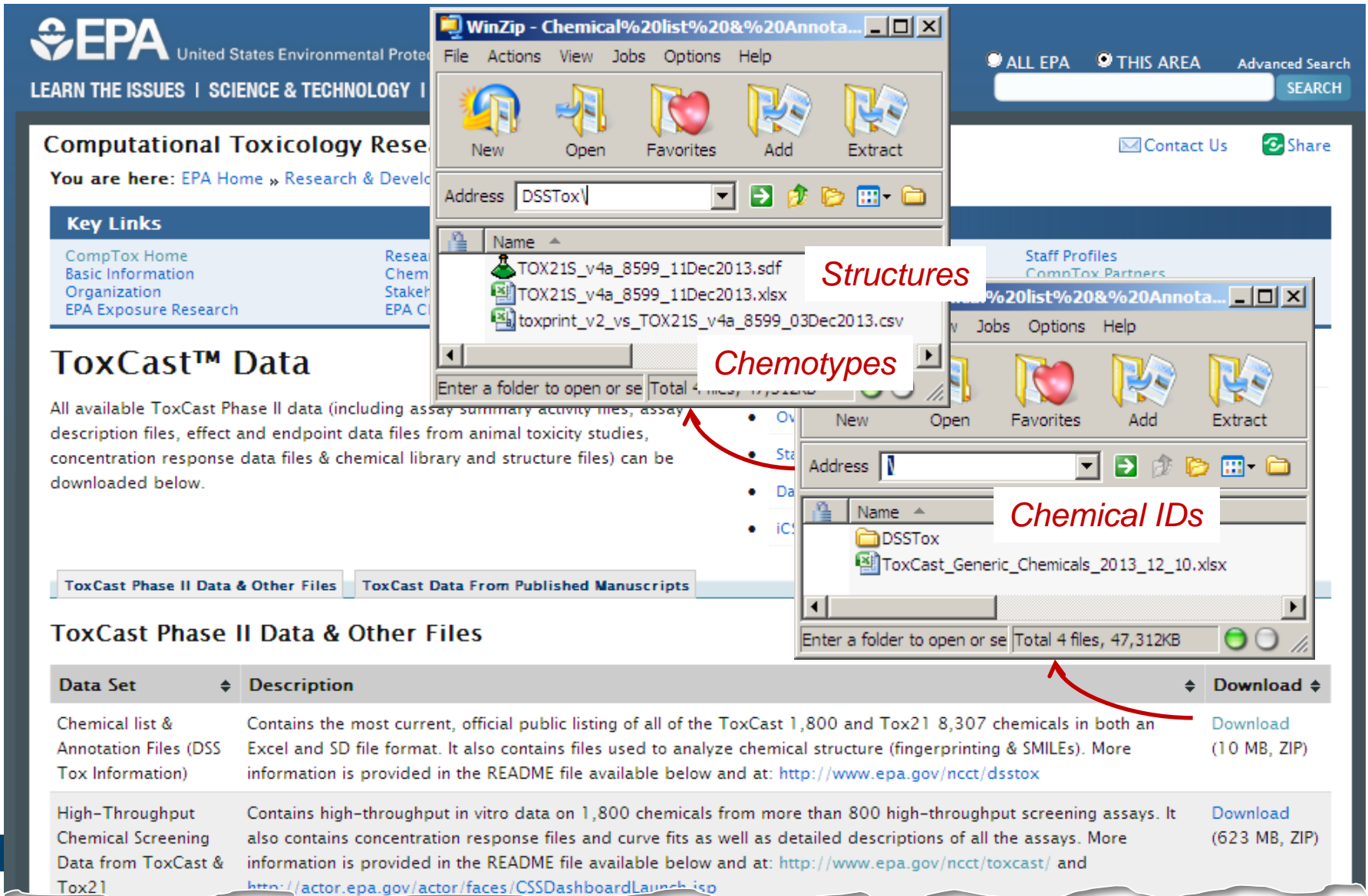
<http://www.epa.gov/ncct/toxcast/>

---

- ToxCast Assay Summary Activity Files
- ToxCast Assay Annotation Files
- **ToxCast Chemical Library & Structure Files (DSSTox)**
- ToxCast Concentration Response Data Files
- ToxRefDB Effect & Endpoint Data Files

# ToxCast Data: Chemical Files

<http://www.epa.gov/ncct/toxcast/data.html>



The screenshot shows the EPA ToxCast Data website with two WinZip file explorer windows overlaid. The top window shows a folder named 'DSSTox' containing files: 'TOX21S\_v4a\_8599\_11Dec2013.sdf', 'TOX21S\_v4a\_8599\_11Dec2013.xlsx', and 'toxprint\_v2\_vs\_TOX21S\_v4a\_8599\_03Dec2013.csv'. A red arrow points from the 'Chemotypes' label to the 'TOX21S\_v4a\_8599\_11Dec2013.sdf' file. The bottom window shows a folder named 'DSSTox' containing files: 'DSSTox' (folder) and 'ToxCast\_Generic\_Chemicals\_2013\_12\_10.xlsx'. A red arrow points from the 'Chemical IDs' label to the 'ToxCast\_Generic\_Chemicals\_2013\_12\_10.xlsx' file. The website background includes a navigation bar, a search bar, and a table of data sets.

**Structures**

**Chemotypes**

**Chemical IDs**

**ToxCast™ Data**

All available ToxCast Phase II data (including assay summary activity files, assay description files, effect and endpoint data files from animal toxicity studies, concentration response data files & chemical library and structure files) can be downloaded below.

**ToxCast Phase II Data & Other Files** | **ToxCast Data From Published Manuscripts**

Data Set	Description	Download
Chemical list & Annotation Files (DSS Tox Information)	Contains the most current, official public listing of all of the ToxCast 1,800 and Tox21 8,307 chemicals in both an Excel and SD file format. It also contains files used to analyze chemical structure (fingerprinting & SMILEs). More information is provided in the README file available below and at: <a href="http://www.epa.gov/ncct/dsstox">http://www.epa.gov/ncct/dsstox</a>	<a href="#">Download</a> (10 MB, ZIP)
High-Throughput Chemical Screening Data from ToxCast & Tox21	Contains high-throughput in vitro data on 1,800 chemicals from more than 800 high-throughput screening assays. It also contains concentration response files and curve fits as well as detailed descriptions of all the assays. More information is provided in the README file available below and at: <a href="http://www.epa.gov/ncct/toxcast/">http://www.epa.gov/ncct/toxcast/</a> and <a href="http://actor.epa.gov/actor/faces/CSSDashboardLaunch.jsp">http://actor.epa.gov/actor/faces/CSSDashboardLaunch.jsp</a>	<a href="#">Download</a> (623 MB, ZIP)

## Computational Toxicology Research

You are here: [EPA Home](#) » [Research & Development](#) » [CompTox](#) » [ToxCast™](#)

### Key Links

<a href="#">CompTox Home</a>	<a href="#">Research Projects</a>
<a href="#">Basic Information</a>	<a href="#">Chemical Databases</a>
<a href="#">Organization</a>	<a href="#">Stakeholder Workshops</a>
<a href="#">EPA Exposure Research</a>	<a href="#">EPA Chemical Safety Research</a>

## ToxCast™ Data

All available ToxCast Phase II data (including assay summary activity files, assay description files, effect and endpoint data files from animal toxicity studies, concentration response data files & chemical library and structure files) can be downloaded below.

[ToxCast Phase II Data & Other Files](#) [ToxCast Data From Published Manuscripts](#)

### ToxCast Phase II Data & Other Files

Data Set	Description	
Chemical list & Annotation Files (DSS Tox Information)	Contains the most current, official public listing of all of the chemicals in the ToxCast Phase II data. It also contains files used to analyze chemical structure (fingerprinting & SMILES). More information is provided in the README file available below and at: <a href="http://www.epa.gov/ncct/dssto">http://www.epa.gov/ncct/dssto</a>	(10 MB, ZIP)

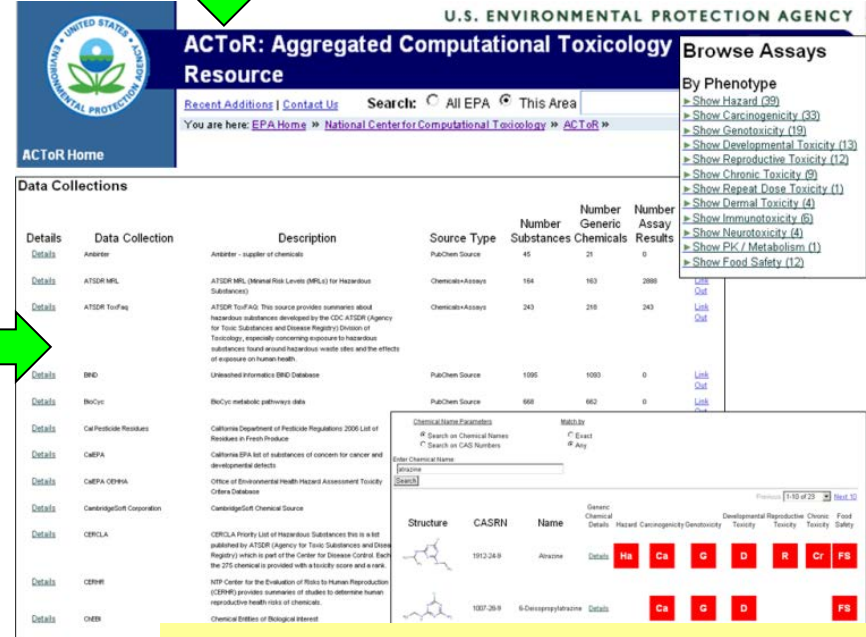
### Contents

- Introduction1 .....
- How were chemicals selected? .....
- How are chemicals identified? .....
- How are chemicals being QC'd? .....
- What assays were run and where were they run? .....
- What does the assay data look like? .....
- How was the assay data processed? .....
- How is the assay data being QC'd? .....
- What information is provided for *in vivo* animal tests in ToxRes? .....
- What other information are we providing in this data release? .....
- Appendix 1: ToxCast *In Vitro* Data Processing Pipeline.....
- Appendix 2: Assay Description files .....
- Appendix 3: Chemical Information .....
- Appendix 4: Chemical Selection, Processing and QC .....

ToxCast Phase II Data Release README File This file is intended to be a brief guide to help navigate and analyze all the ToxCast Phase II data.

[Download](#)  
(481 KB, PDF)

- Source of high quality structure files and chemical IDs for Tox21 (TOX21S) and ToxCast (TOXCST) programs

**U.S. ENVIRONMENTAL PROTECTION AGENCY**  
**ACToR: Aggregated Computational Toxicology Resource**

Recent Additions | Contact Us | Search:  All EPA  This Area

You are here: EPA Home » National Center for Computational Toxicology » ACToR »

**ACToR Home**

**Browse Assays**

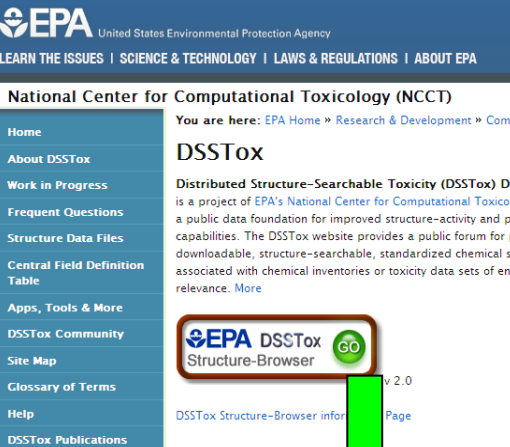
By Phenotype

- Show Hazard (39)
- Show Carcinogenicity (33)
- Show Genotoxicity (19)
- Show Developmental Toxicity (13)
- Show Reproductive Toxicity (12)
- Show Chronic Toxicity (9)
- Show Repeat Dose Toxicity (1)
- Show Dermal Toxicity (4)
- Show Immunotoxicity (6)
- Show Neurotoxicity (4)
- Show PK / Metabolism (1)
- Show Food Safety (12)

**Data Collections**

Details	Data Collection	Description	Source Type	Number Substances	Number Generic Chemicals	Number Assay Results
Details	Antider	Antider - supplier of chemicals	PubChem Source	45	21	0
Details	ATSDR MEL	ATSDR MEL (Minimal Risk Levels (MRLs) for Hazardous Substances)	Chemicals+Assays	164	163	2880
Details	ATSDR ToxFAQ	ATSDR ToxFAQ. This source provides summaries about hazardous substances developed by the CDC ATSDR (Agency for Toxic Substances and Disease Registry) Division of Toxicology, especially concerning exposures to hazardous substances found around hazardous waste sites and the effects of exposure on human health.	Chemicals+Assays	243	216	243
Details	BMD	Unrated Informative BMD Database	PubChem Source	1095	1093	0
Details	BioCyc	BioCyc metabolic pathways data	PubChem Source	668	662	0
Details	Cal Pesticide Residues	California Department of Pesticide Regulations 2006 List of Residues in Fresh Produce				
Details	CaEPA	California EPA list of substances of concern for cancer and developmental effects				
Details	CaEPA-CDRHA	Office of Environmental Health Hazard Assessment Toxicity Criteria Database				
Details	CambridgeIsot Corporation	CambridgeIsot Chemical Source				
Details	CDRLA	CDRLA Priority List of Hazardous Substances this is a list published by ATSDR (Agency for Toxic Substances and Disease Registry) which is part of the Center for Disease Control. Each of the 375 chemicals is provided with a toxicity score and a rank.				
Details	CDRH	NIEHS Center for the Evaluation of Risks to Human Reproduction (CEHR) provides summaries of studies to determine human reproductive health risks of chemicals.				
Details	CMER	Chemical Entities of Biological Interest				

- What data are available across EPA and Internet resources for my chemical or its analogs?



**EPA** United States Environmental Protection Agency

LEARN THE ISSUES | SCIENCE & TECHNOLOGY | LAWS & REGULATIONS | ABOUT EPA

**National Center for Computational Toxicology (NCCT)**

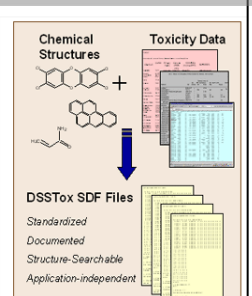
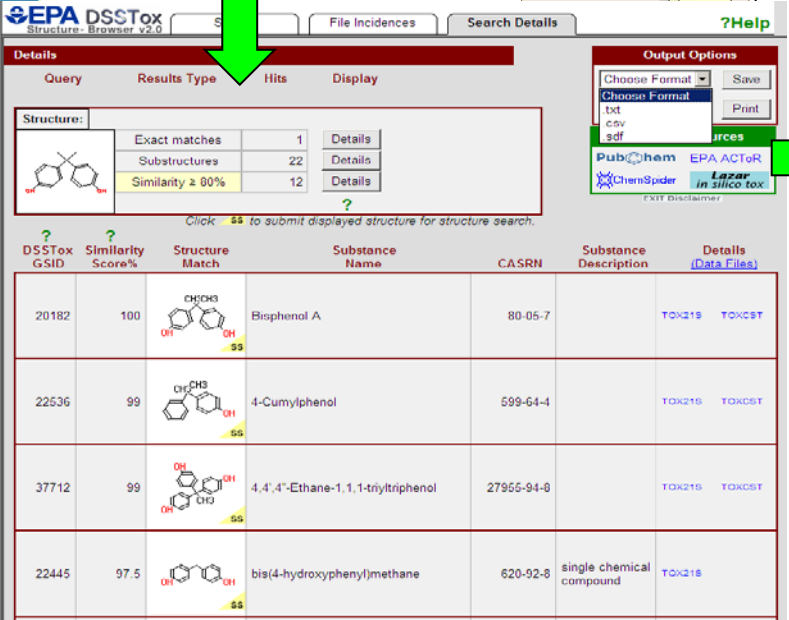
You are here: EPA Home » Research & Development » Computational Toxicology » DSSTox

**DSSTox**

Distributed Structure-Searchable Toxicity (DSSTox) Database Network is a project of EPA's National Center for Computational Toxicology, helping to build a public data foundation for improved structure-activity and predictive toxicology capabilities. The DSSTox website provides a public forum for publishing downloadable, structure-searchable, standardized chemical structure files associated with chemical inventories or toxicity data sets of environmental relevance. More

**EPA DSSTox Structure-Browser** GO

DSSTox Structure-Browser info Page


**EPA DSSTox Structure-Browser v2.0**


File Incidences Search Details ?Help





**Details**

Query Results Type Hits Display

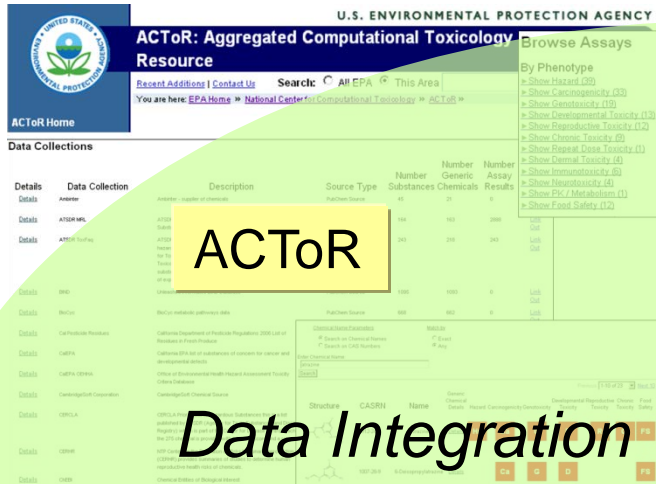
Structure:

Structure	Exact matches	Substructures	Similarity ≥ 80%
	1	22	12

Click  to submit displayed structure for structure search.

DSSTox GSID	Similarity Score%	Structure Match	Substance Name	CASRN	Substance Description	Details (Data Files)
20182	100		Bisphenol A	80-05-7		TOX21S TOXCST
22536	99		4-Cumylphenol	599-64-4		TOX21S TOXCST
37712	99		4,4'-(4-Ethane-1,1,1-triyl)triphenol	27955-94-8		TOX21S TOXCST
22445	97.5		bis(4-hydroxyphenyl)methane	620-92-8	single chemical compound	TOX21S
47962						

- Is my chemical in ToxCast or Tox21?
- Are "similar" chemicals being tested?



**U.S. ENVIRONMENTAL PROTECTION AGENCY**  
**ACToR: Aggregated Computational Toxicology Resource**

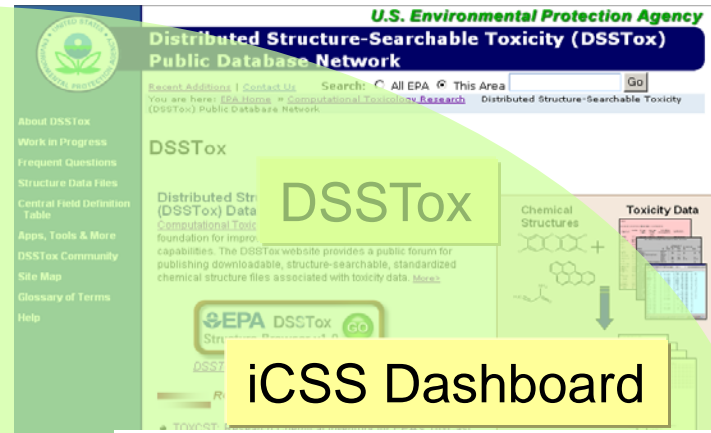
Recent Address | Contact Us | Search: All EPA | This Area | You are here: EPA Home > National Center for Computational Toxicology > ACToR

**ACToR Home**

**Data Collections**

Details	Data Collection	Description	Source Type	Number of Substances	Number of Generic Chemicals	Number of Assays	Number of Results
Details	ATISER MRL	ATISER MRL	ATISER MRL	140	140	140	140
Details	ATISER MRL	ATISER MRL	ATISER MRL	140	140	140	140

**ACToR**



**U.S. Environmental Protection Agency**  
**Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network**

Recent Address | Contact Us | Search: All EPA | This Area | Go | You are here: EPA Home > Computational Toxicology > Research > Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network.

**DSSTox**

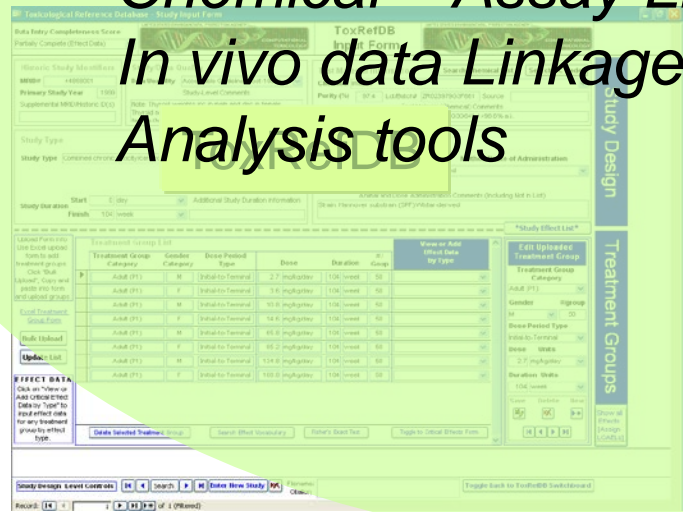
Distributed Structure-Searchable Toxicity (DSSTox) Data Computational Toxicology Foundation for improved capabilities. The DSSTox website provides a public forum for publishing downloadable, structure-searchable, standardized chemical structure files associated with toxicity data.

**DSSTox**

**iCSS Dashboard**

*Data Integration*  
*Chemical – Assay Linkage*  
*In vivo data Linkage*  
*Analysis tools*

<http://actor.epa.gov/dashboard/>



**Technological Reference Database - Study List**

Partial Complete (Effect Data)

Search: Study Identifiers  
Priority Study Year: 1990  
Supplemental MRL Product: 0.01

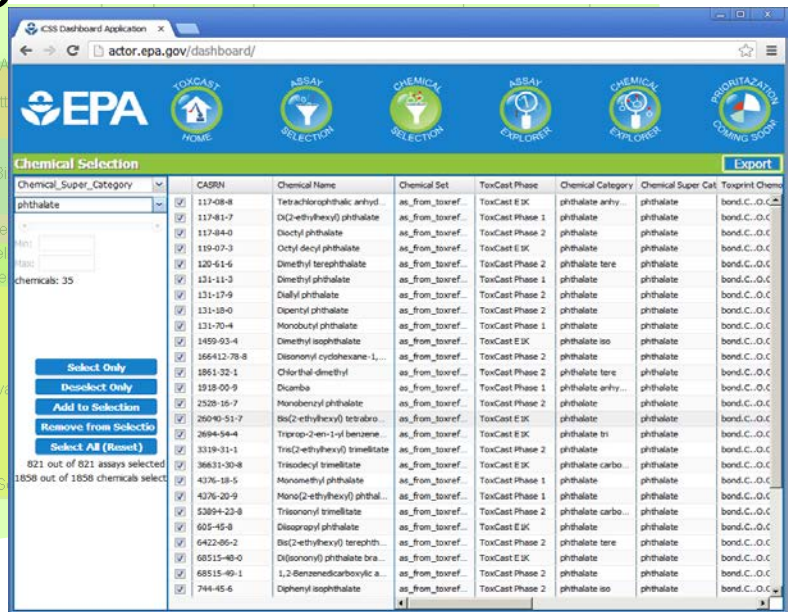
Study Type: Control Group

Study Duration: Start: 1/1/2000, End: 1/1/2000

Treatment Group Category	Gender	Dose	Duration	Effect	Group
AA04 (01)	M	Initial to Terminal	3.7 mg/kg/day	100% survival	30
AA04 (01)	F	Initial to Terminal	3.7 mg/kg/day	100% survival	30
AA04 (01)	M	Initial to Terminal	14.8 mg/kg/day	100% survival	30
AA04 (01)	F	Initial to Terminal	14.8 mg/kg/day	100% survival	30
AA04 (01)	M	Initial to Terminal	60.0 mg/kg/day	100% survival	30
AA04 (01)	F	Initial to Terminal	60.0 mg/kg/day	100% survival	30
AA04 (01)	M	Initial to Terminal	124.0 mg/kg/day	100% survival	30
AA04 (01)	F	Initial to Terminal	124.0 mg/kg/day	100% survival	30

**EFFECT DATA**

821 out of 821 assays selected  
1858 out of 1858 chemicals selected



**iCSS Dashboard Application**

[actor.epa.gov/dashboard/](http://actor.epa.gov/dashboard/)

**Chemical Selection**

Chemical_Super_Category	CASRN	Chemical Name	Chemical Set	ToxCast Phase	Chemical Category	Chemical Super Cat.	Toxprint Chem
phthalate	117-08-8	Tetrachlorophthalic anhyd.	as_from_isoref...	ToxCast E 1K	phthalate anhy...	phthalate	bond.C..O.C...
	117-81-7	Di(2-ethylhexyl) phthalate	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	117-84-0	Diethyl phthalate	as_from_isoref...	ToxCast Phase 2	phthalate	phthalate	bond.C..O.C...
	119-07-3	Octyl decyl phthalate	as_from_isoref...	ToxCast E 1K	phthalate	phthalate	bond.C..O.C...
	120-61-6	Dimethyl terephthalate	as_from_isoref...	ToxCast Phase 2	phthalate tere	phthalate	bond.C..O.C...
	131-14-3	Dimethyl phthalate	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	131-17-9	Diethyl phthalate	as_from_isoref...	ToxCast Phase 2	phthalate	phthalate	bond.C..O.C...
	131-19-0	Dioctyl phthalate	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	131-70-4	Monobutyl phthalate	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	1459-03-4	Dimethyl isophthalate	as_from_isoref...	ToxCast E 1K	phthalate iso	phthalate	bond.C..O.C...
	1664-12-8	Dibutyl sebacate	as_from_isoref...	ToxCast Phase 2	phthalate	phthalate	bond.C..O.C...
	1861-32-1	Chlorthal-dimethyl	as_from_isoref...	ToxCast Phase 2	phthalate tere	phthalate	bond.C..O.C...
	1918-00-9	Diamba	as_from_isoref...	ToxCast Phase 1	phthalate anhy...	phthalate	bond.C..O.C...
	2520-16-7	Monobenzyl phthalate	as_from_isoref...	ToxCast Phase 2	phthalate	phthalate	bond.C..O.C...
	26040-51-7	Bis(2-ethylhexyl) tetrabro...	as_from_isoref...	ToxCast E 1K	phthalate	phthalate	bond.C..O.C...
	2694-94-4	Triprop-2-en-1-yl benzene	as_from_isoref...	ToxCast E 1K	phthalate tri	phthalate	bond.C..O.C...
	3319-31-1	Tri(2-ethylhexyl) trimellate	as_from_isoref...	ToxCast Phase 2	phthalate tere	phthalate	bond.C..O.C...
	36631-30-8	Tridecyl trimellate	as_from_isoref...	ToxCast E 1K	phthalate carbo...	phthalate	bond.C..O.C...
	4376-18-5	Monomethyl phthalate	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	4376-20-9	Mono(2-ethylhexyl) phthal...	as_from_isoref...	ToxCast Phase 1	phthalate	phthalate	bond.C..O.C...
	5304-23-8	Triisononyl trimellate	as_from_isoref...	ToxCast Phase 1	phthalate carbo...	phthalate	bond.C..O.C...
	605-45-8	Diisopropyl phthalate	as_from_isoref...	ToxCast E 1K	phthalate	phthalate	bond.C..O.C...
	6422-86-2	Bis(2-ethylhexyl) terephth...	as_from_isoref...	ToxCast Phase 2	phthalate tere	phthalate	bond.C..O.C...
	68515-48-0	Di(isononyl) phthalate bra...	as_from_isoref...	ToxCast E 1K	phthalate	phthalate	bond.C..O.C...
	68515-49-1	1,2-Benzenedicarboxylic a...	as_from_isoref...	ToxCast Phase 2	phthalate	phthalate	bond.C..O.C...
	744-45-6	Diphenyl isophthalate	as_from_isoref...	ToxCast Phase 2	phthalate iso	phthalate	bond.C..O.C...

# ToxCast Inventories

---

**293** • ToxCast Phase I (293 unique cmpds)

- *EPA pesticidal actives w/ rich in vivo data*
- *PFOAs, BPA, approx 12 metabolite/parent pairs*

**1060** • ToxCast Phase II (767 unique new cmpds)

- *EPA pesticides, high interest EPA and stakeholder inventories, data rich chemicals (EDSP, OPPT, antimicrobials, inerts, green alternatives, fragrances, water ...)*
- *FDA CFSAN data rich, NCTR LTKB Priority 1 drugs*
- *Toxicity reference chemicals, data-rich chemicals, NTP immunotox*
- *135 Donated pharma cmpds -- failed drugs w/ pre-clinical or clinical tox data*

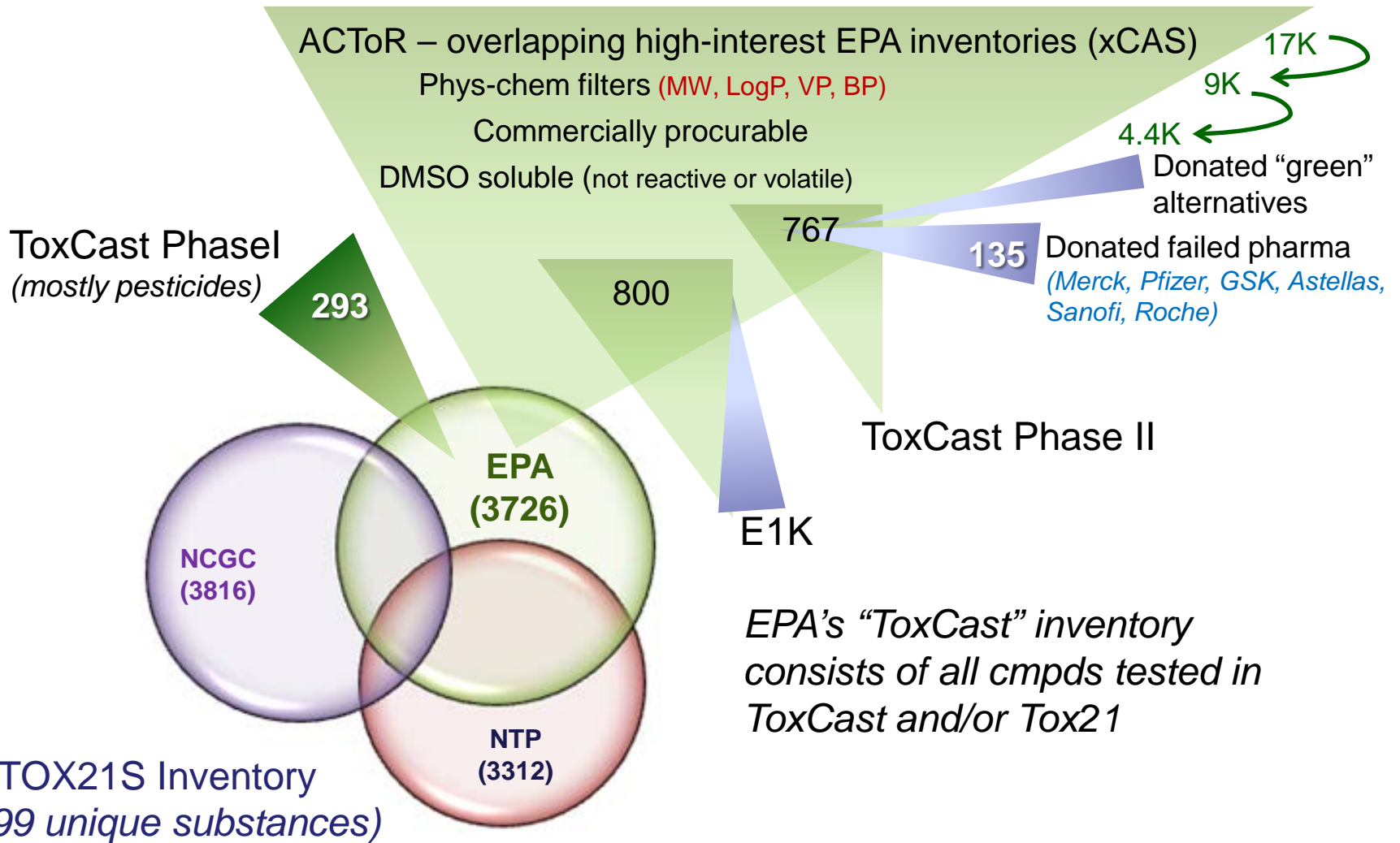
**1860** • ToxCast E1K (800 unique new cmpds)

- *Endocrine active reference cmpds, SAR predicted ER-active/inactives, EDSP cmpds*

**3727** • EPA's Tox21 library (3727 unique cmpds out of current 8599 total)






- *Complete on-hand EPA sample library used to build ToxCast inventories*

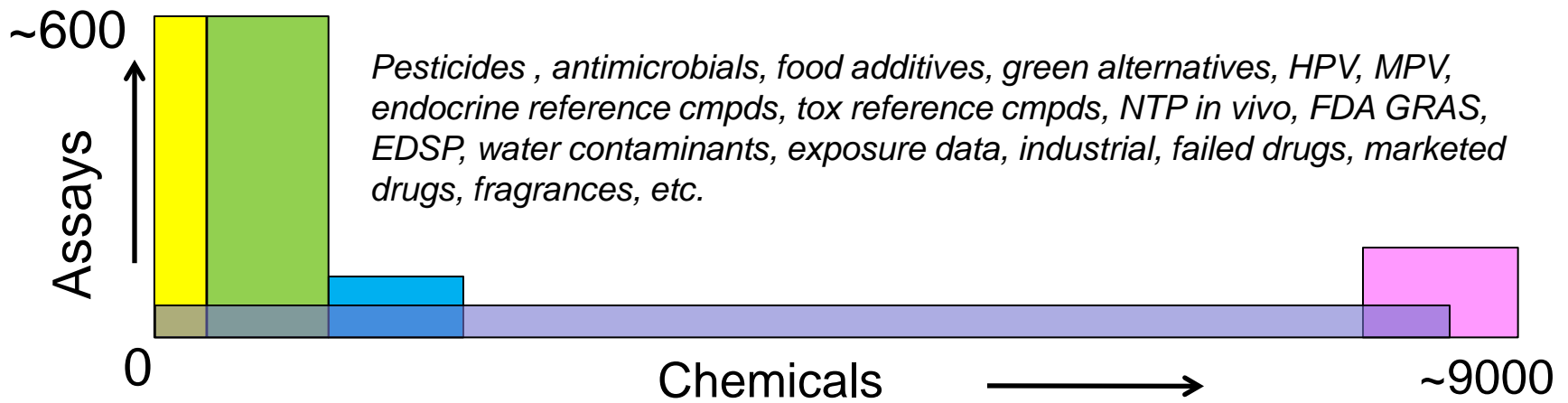
# EPA ToxCast Inventories



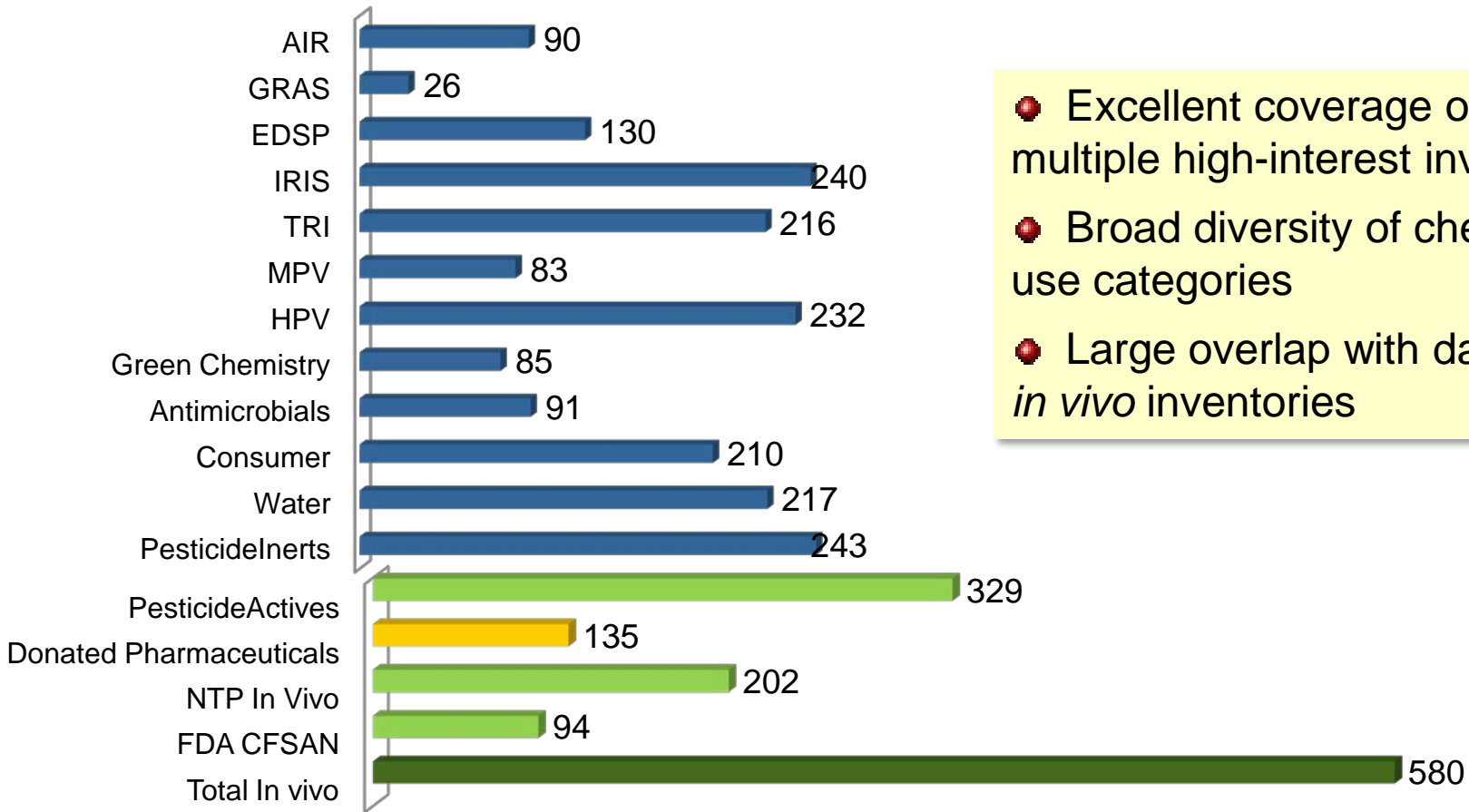


# ToxCast & Tox21: Chemicals, Data and Release Timelines

Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	 293	~600	~700	2011	Now
ToxCast Phase II	 767	~600	~700	03/2013	Now
ToxCast Phase III	 ~900	~100	~100	Just starting	2014
E1K (endocrine)	 800	~50	~120	03/2013	Now
Tox21	 8,599	~25	~50	Ongoing	Ongoing



# ToxCast PhI&PhII 1060: # Compounds per Inventory



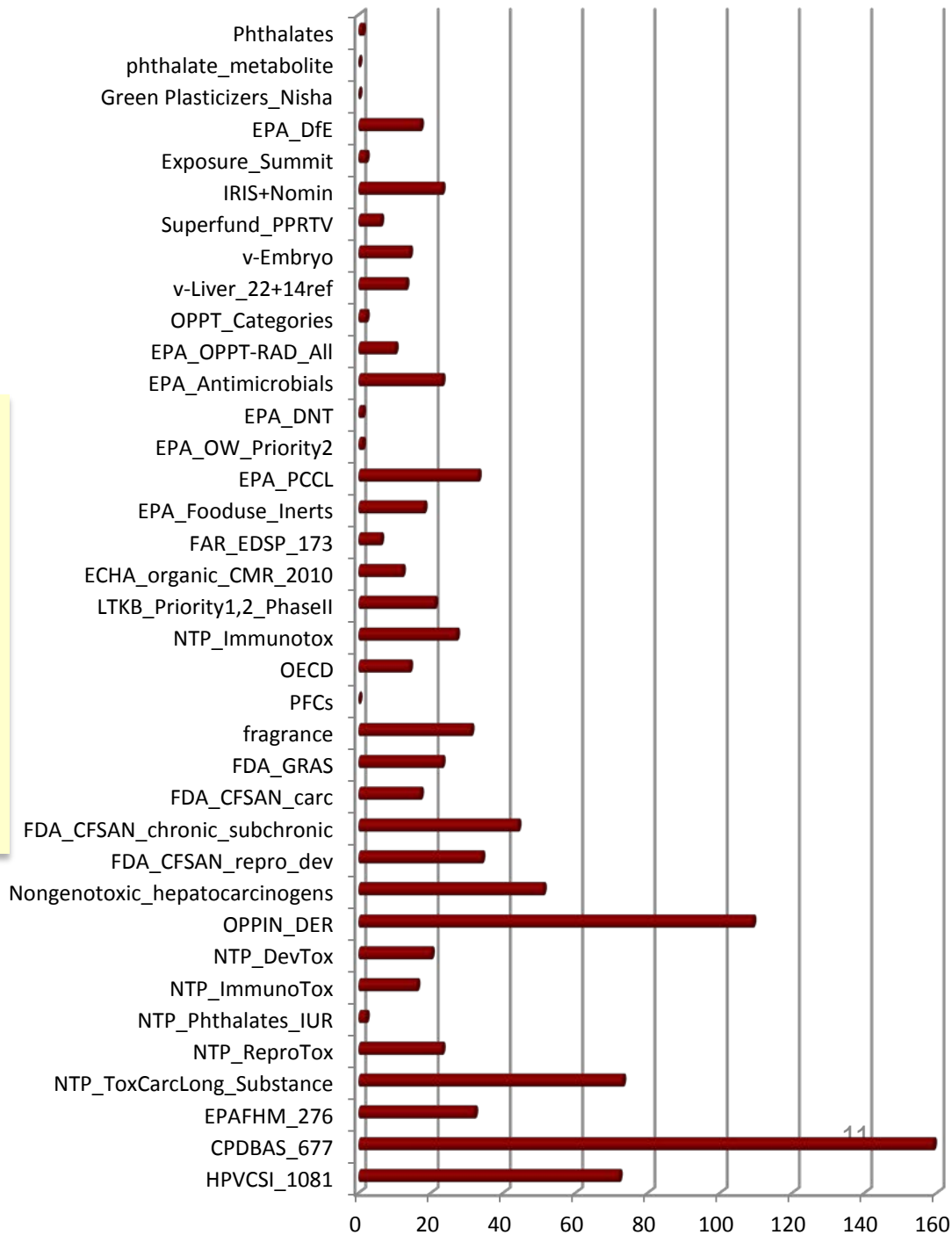
- Excellent coverage of multiple high-interest inventories
- Broad diversity of chemical-use categories
- Large overlap with data-rich *in vivo* inventories

# ToxCast Phase II Pharmaceuticals: *Multipurposing*

• 275 PhII cmpds classified as drugs based on presence on FDAMDD, NCGC, NCTR LKB, Donated pharma lists

• 150 drugs appear on additional lists (i.e., multipurposing)

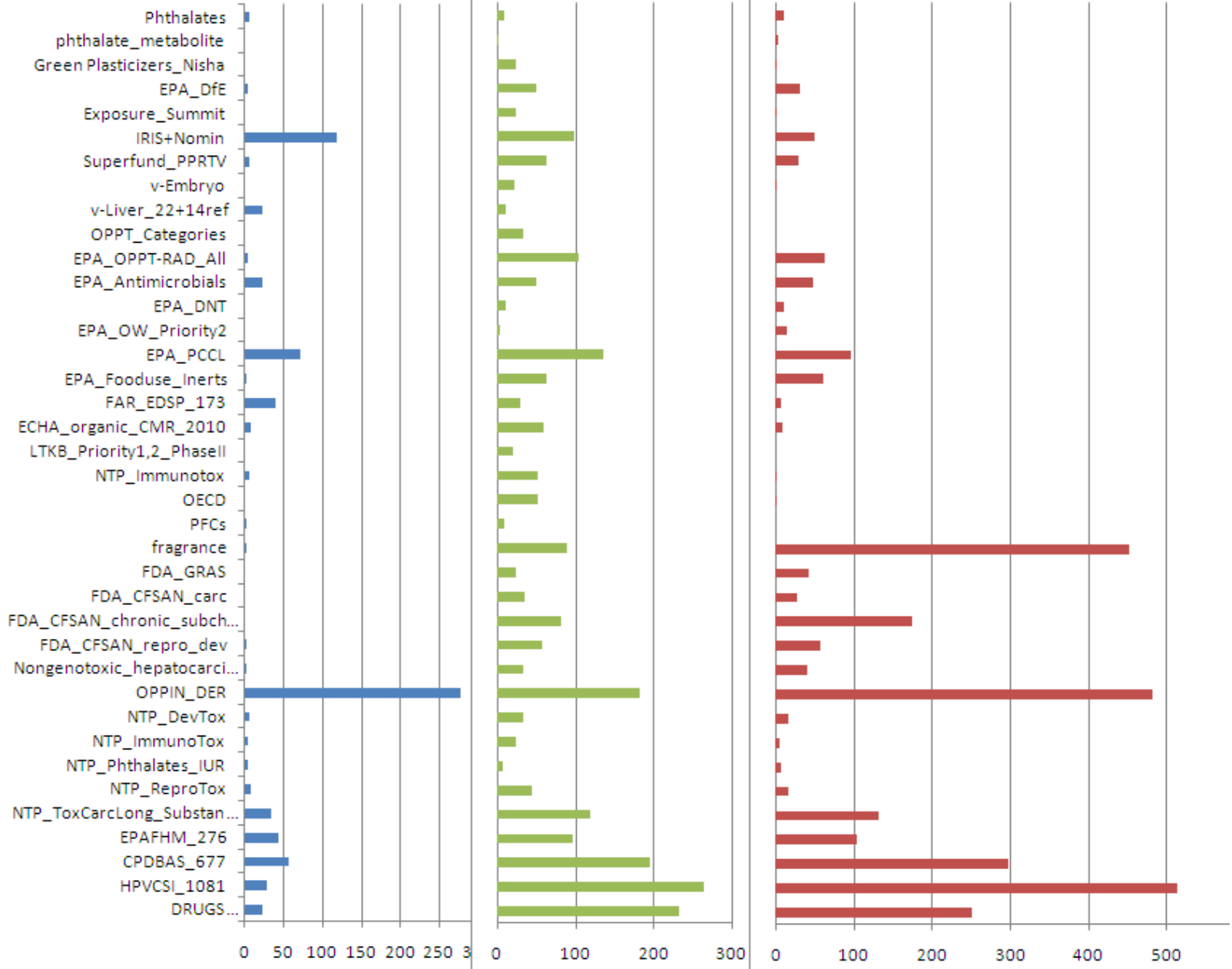
• Caffeine appears on 18 lists



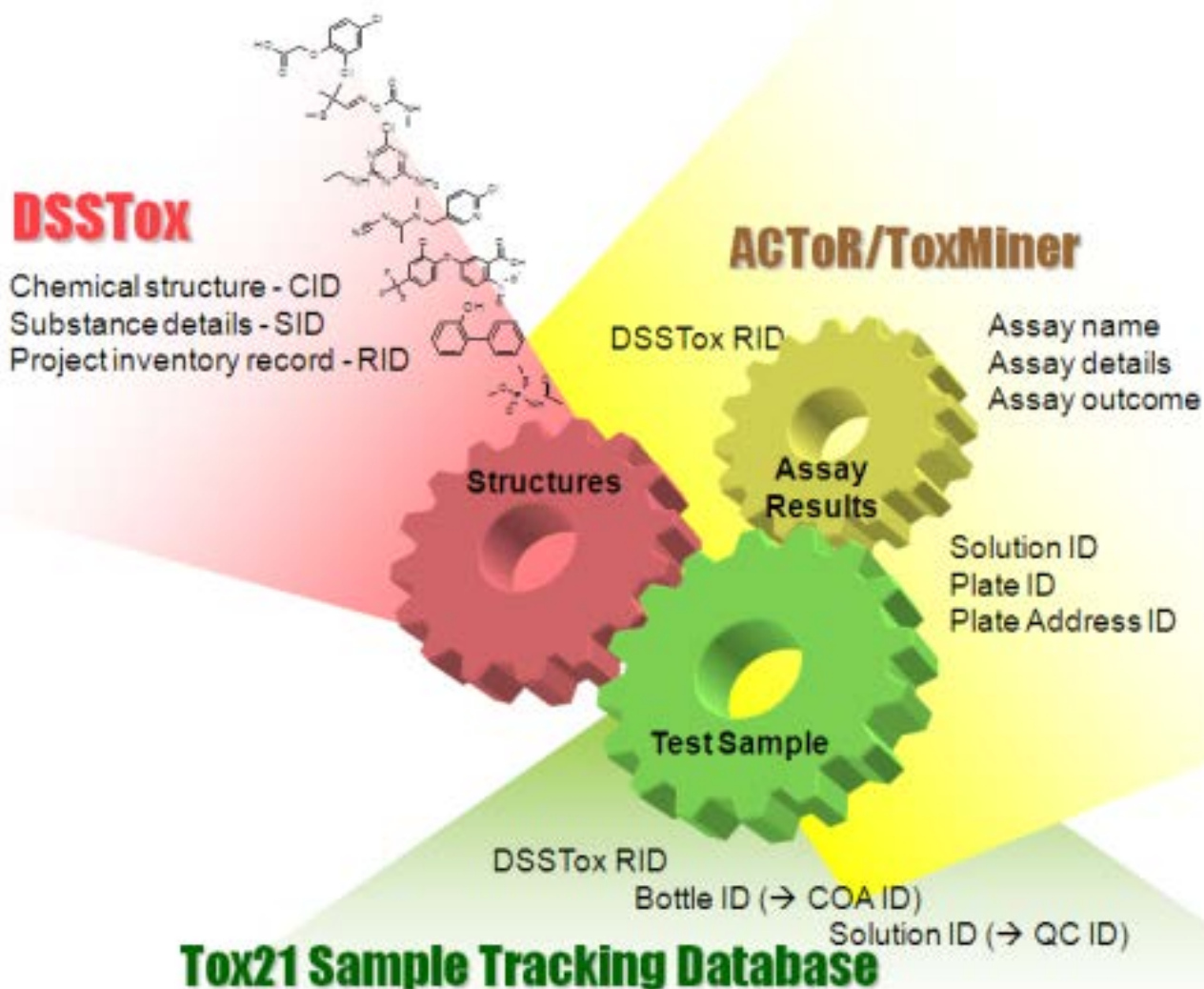
ToxCastPhI\_293

ToxCastPhII\_767

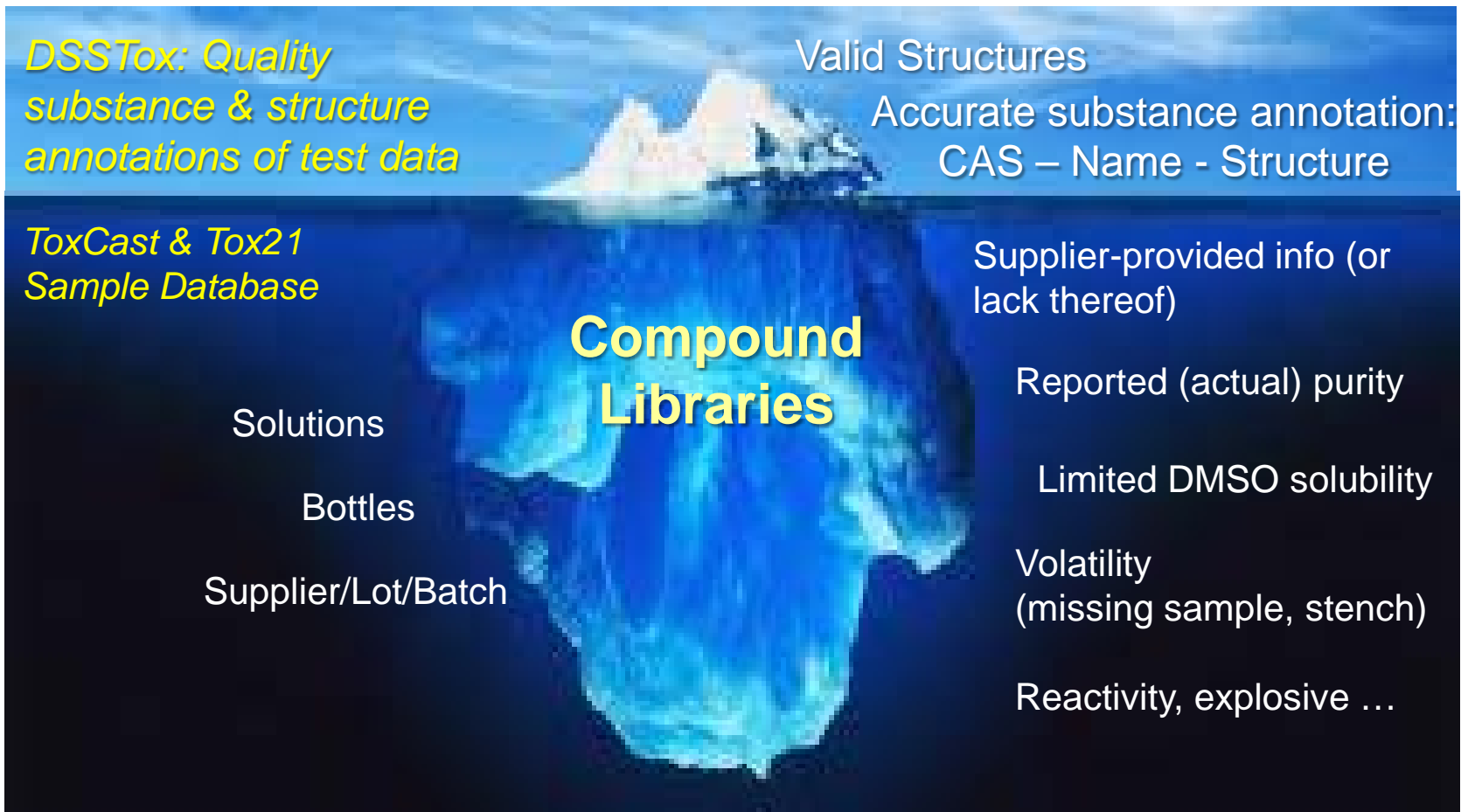
Tox21-EPA\_2659



# Integrated Chemical-Assay Data Management



# Chemical & Data Quality Issues



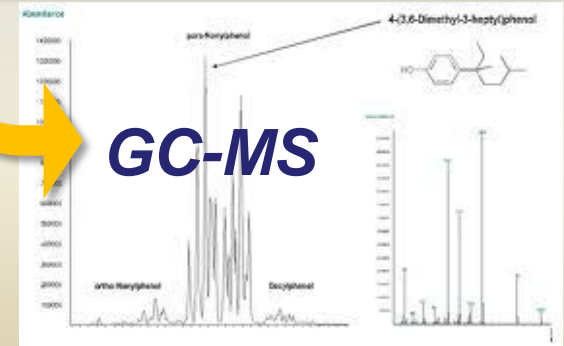


A copy of each parent Tox21 assay plate is subjected to analytical QC for assessing purity, identity, concentration, stability

PASS = Confirm parent ion peak and >90% purity

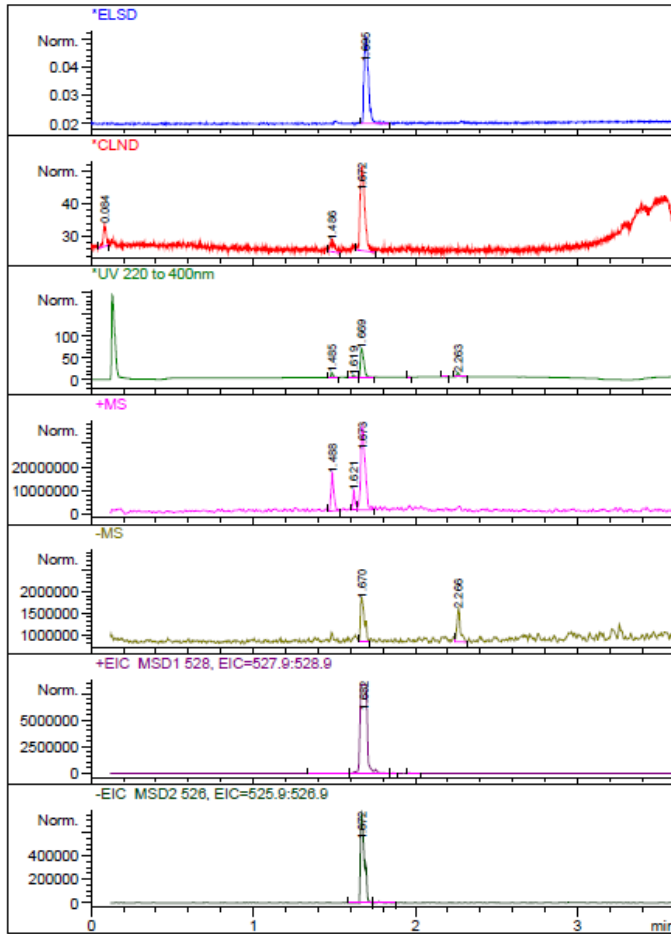
Fail, inconclusive or analytical method inappropriate

Retest at later time point under assay conditions for stability

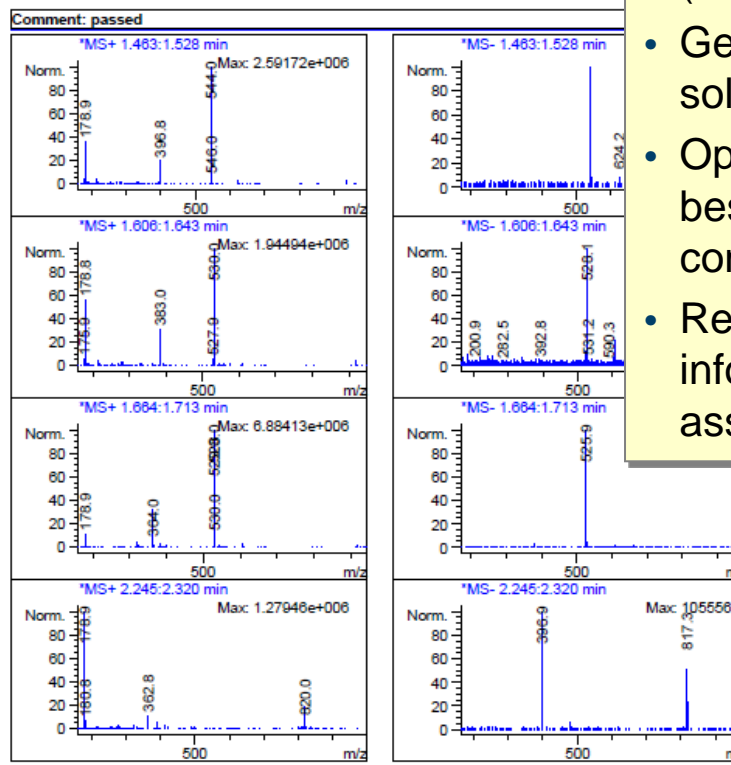


Publish QC summary results in association with assay data

ID Tox21\_110896 Plate FDA-plate1 Well P1-4-03 File .....00024246-05-D Inj Date: 30 Jan 10 7:37 am - MF C27H29NO10 MW 527.2 Expected Conc: 3.00 mM



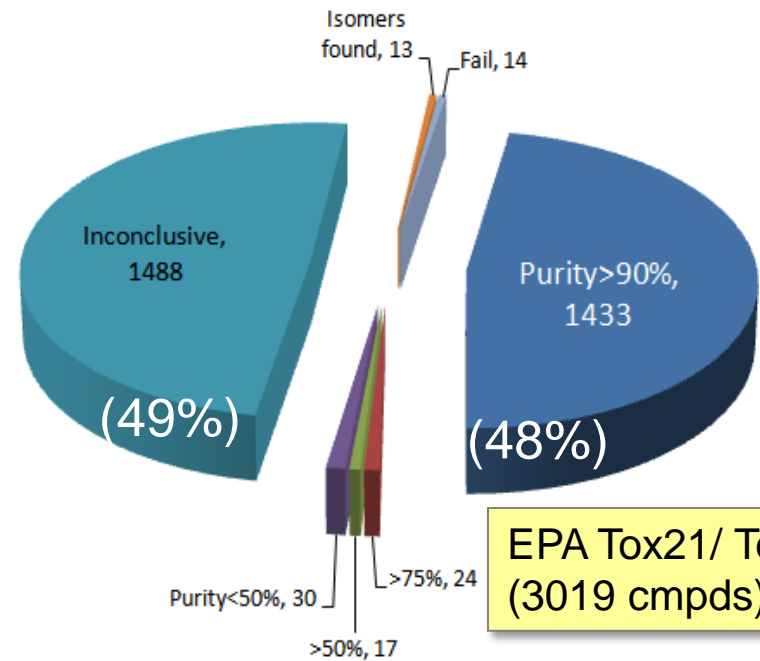
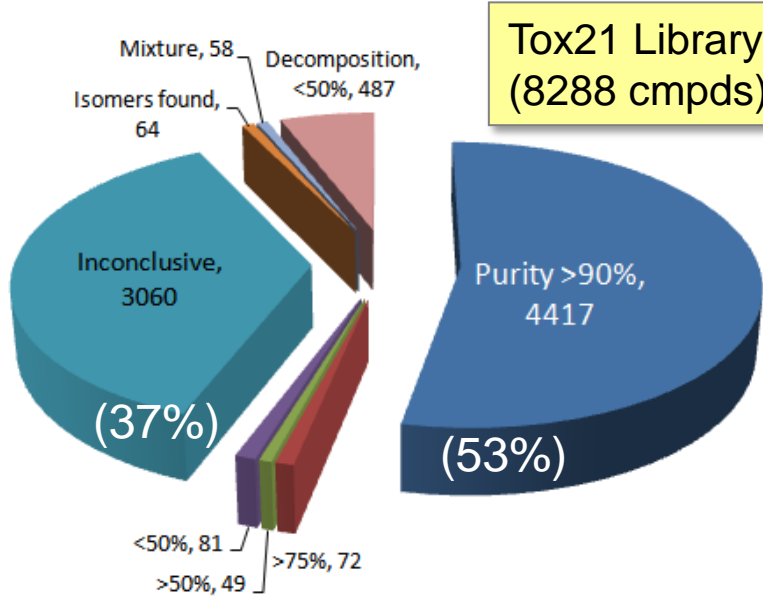
RT	Found	ELS%	UV %	ELS [mg/mL]	Adj [ELS]	[N mM]	Adj [CLN]	#N
0.08		0.0	0.0			0.27 mM		
1.48		0.0	8.2			0.24 mM		
1.62		0.0	3.7					
1.67	Yes	100.0	77.7	1.5	2.85 mM	1.71 mM	1.71	
1.96		0.0	0.4					
2.17		0.0	1.3					
2.26		0.0	8.7					



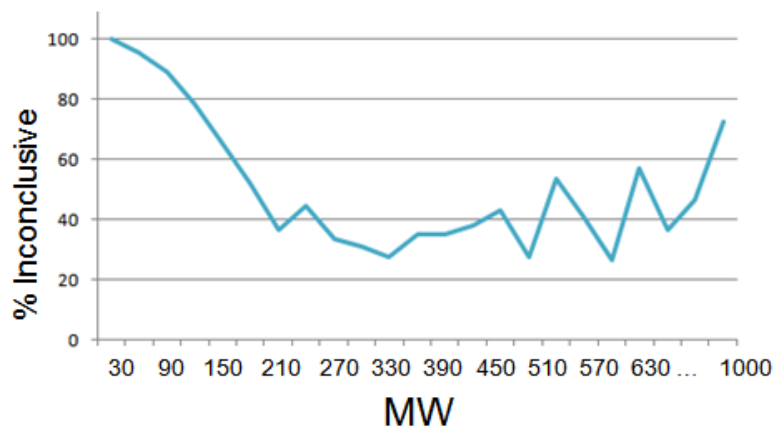
- Analytical QC performed on entire Tox21 chemical library at soln-well-level (i.e., test plate copy x time)
- Generate summary calls at soln and compd level
- Open discussion on how best to release and communicate QC results
- Recommend using to inform interpretation of assay results



**Tox21 Library  
(8288 cmpds)**

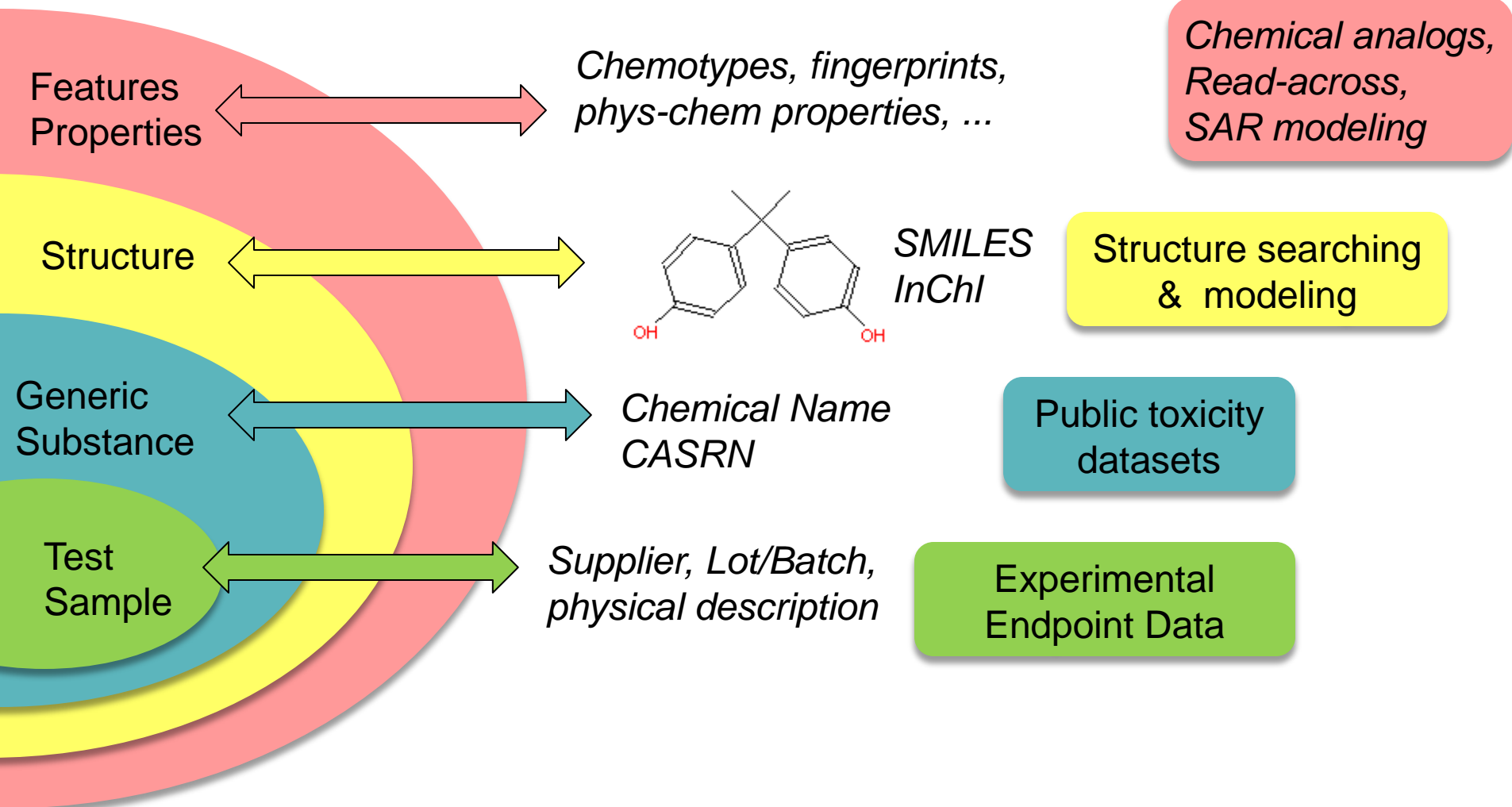


**EPA Tox21/ ToxCast  
(3019 cmpds)**



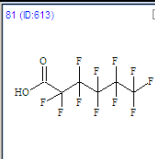
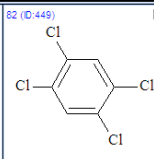
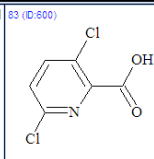
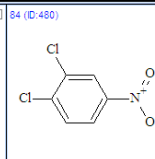
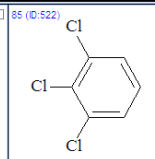
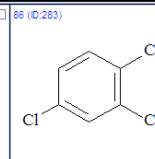
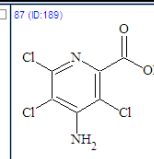
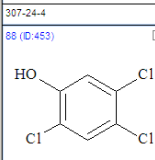
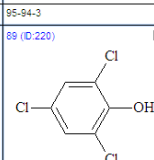
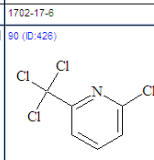
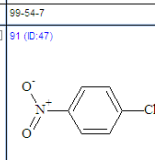
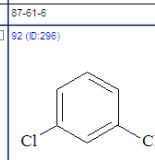
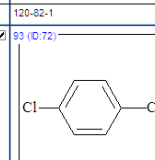
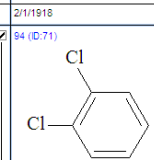
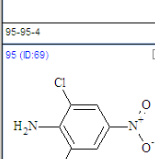
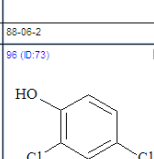
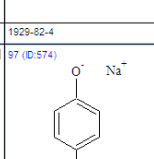
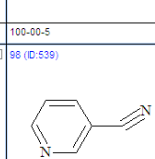
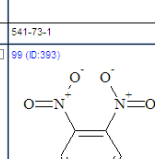
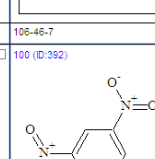
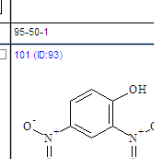
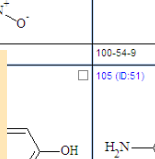
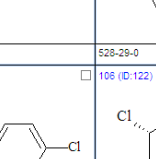
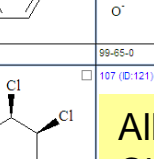
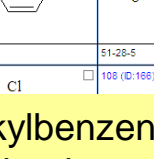
- “inconclusives” require additional follow-up & method development
- EPA’s library contains higher proportion of inconclusives (e.g., low MW cmpds, dyes, metals)
- Overall very low “fail” rate

# Chemical Elements to Data Integration: *Chemical representations* → *Uses*



DSSTox_RID	
DSSTox_GSID	
DSSTox_CID	
DSSTox_FileID	
TestSubstance_ChemicalName	
TestSubstance_CASRN	
TestSubstance_Description	
ChemicalNote	
STRUCTURE_Shown	
STRUCTURE_Formula	
STRUCTURE_MolecularWeight	
STRUCTURE_ChemicalType	
STRUCTURE_TestForm_DefinedOrganic	
STRUCTURE_ChemicalName_IUPAC	
STRUCTURE_SMILES	
STRUCTURE_Parent_SMILES	
STRUCTURE_InChI	
STRUCTURE_InChIKey	
Substance_modify_yyyymmdd	
ToxCast_TestingStatus	
ToxCast_PhI_v1	
ToxCast_PhI_v2	
ToxCast_PhII	
ToxCast_PhII_DonatedPharma	
ToxCast_e1k	
ToxCast_ToX21	
AnalyticalQC_Problems	
Note_TOXCST	

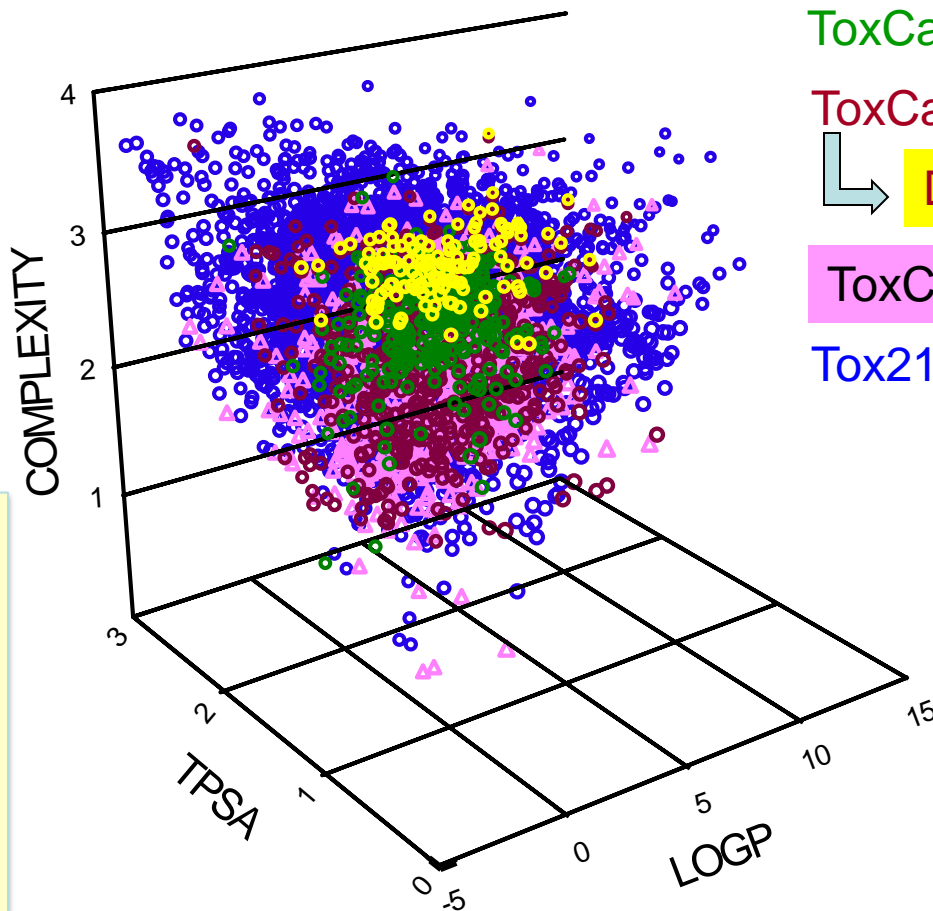
81 (D:613)	82 (D:449)	83 (D:600)	84 (D:480)	85 (D:522)	86 (D:283)	87 (D:189)
						
307-24-4	95-64-3	1702-17-6	99-54-7	87-61-6	120-82-1	2/1/1918
88 (D:453)	89 (D:220)	90 (D:426)	91 (D:47)	92 (D:296)	93 (D:72)	94 (D:71)
						
95-95-4	88-06-2	1929-82-4	100-00-5	541-73-1	106-46-7	95-50-1
95 (D:69)	96 (D:73)	97 (D:574)	98 (D:539)	99 (D:393)	100 (D:392)	101 (D:93)
						
			100-54-9	528-29-0	99-65-0	51-28-5
			105 (D:51)	106 (D:122)	107 (D:121)	108 (D:166)
						

**Total 1860 cmpds:**  
 10 macromolecule  
 96 mixture or formulation  
 17 inorganics  
 43 organometallics  
 84 complexes  
 136 salts

Alkylbenzenes  
 Chlorobenzenes  
 Methyl phenols  
 Nitrobenzenes  
 Aromatic amines  
 Phthalates  
 Perfluorinates...

MW ranges from 30  
 (Formalin) to 1700 (Tannic  
 acid)

# ToxCast & Tox21 Property Space



ToxCast Phase I (293)

ToxCast Phase II (767)



Donated Pharma (135)

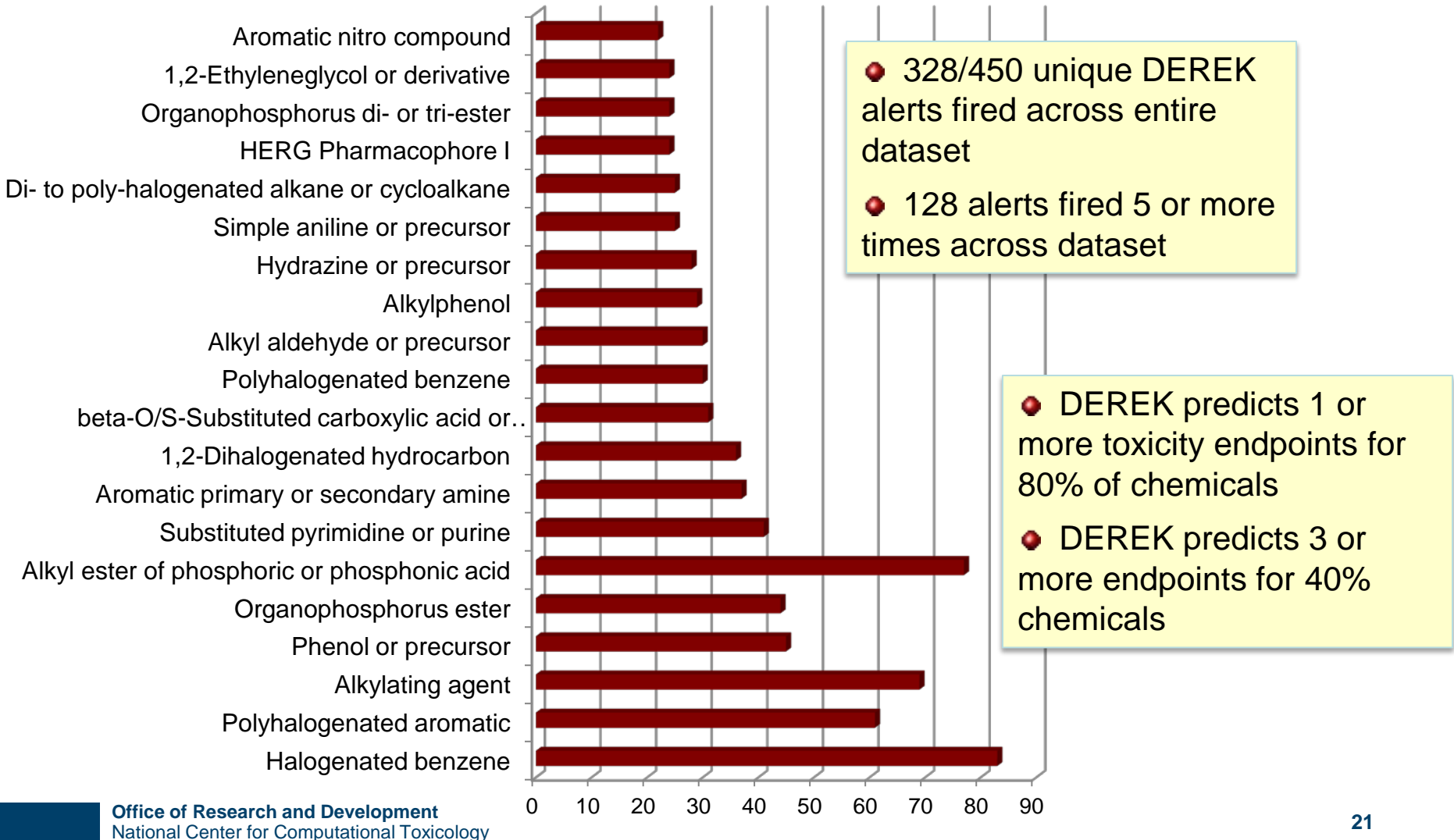
ToxCast e1k (+800)

Tox21 (7324 unique)

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

*Chemical properties computed using "Adrianna" software by Molecular Networks (P. Volarath)*

# Estimating Toxicity Mechanism Coverage: *DEREK (LHASA) Predictions for ToxCast PhII (1060)*



# New Publicly Available Resources: Chemotyper & ToxPrint Chemotypes

Chemotyper: <http://www.chemotyper.org>

- MS Windows application allows for searching and highlighting of chemical chemotypes (chemical substructures or subgraphs) in structure files.
- Developed by *Molecular Networks GmbH (MN)* under contract from FDA CFSAN to house the “ToxPrint” chemotypes

ToxPrint Chemotypes: <http://www.toxprint.org>

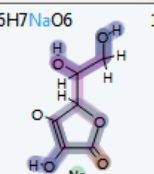
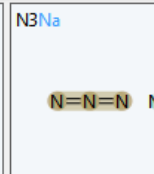
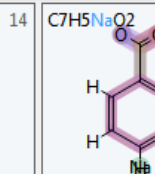
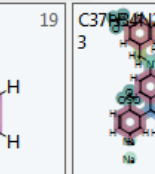
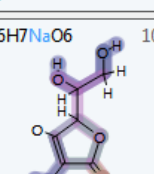
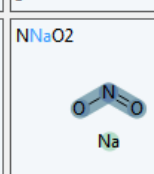
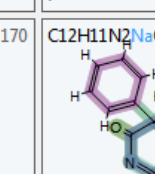
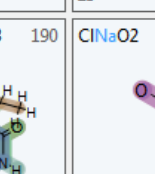
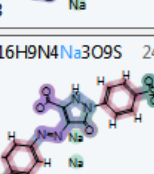
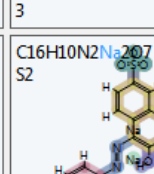
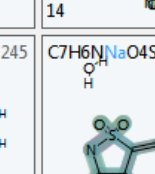
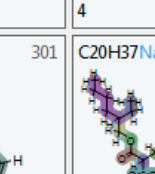
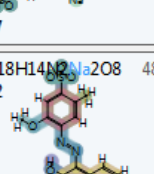
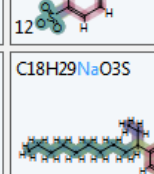
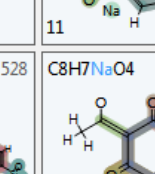
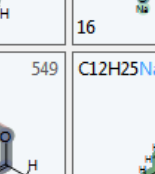
- Developed by *Altamira LLC* for FDA CFSAN’s CERES project
- Designed to provide coverage of EPA & FDA inventories and capture chemical features important for chemical safety assessment workflow
- Contains three subsets:
  - *generic structural fragments (729 total)*
  - *Ashby-Tennant genotoxic carcinogen rules* ([Ashby, J; Tennant, RW, 1988](#))
  - *cancer TTC categories* ([Kroes, R. et. al. 2004](#))



# Chemotyper Application: ToxCast x ToxPrint

## ToxCast PhII (1060)

toxcast\_u1.sdf

 C6H7NaO6 10 13	 N3Na 14 3	 C7H5NaO2 19 7	 C37H44N2Na2O9S 31 23
 C6H7NaO6 106 13	 NNaO2 170 3	 C12H11N2NaO3 190 14	 CINaO2 217 4
 C16H9N4Na3O9S 244 17	 C16H10N2Na2O7 245 12	 C7H6NNaO4S 301 11	 C20H37NaO7S 357 16
 C18H14N2Na2O8 484 17	 C18H29NaO3S 528 16	 C8H7NaO4 549 12	 C12H25NaO4S 550 14

1 / 41

Filter Structures by ID

Filter Chemotypes

Structures Loaded: 1060 Total Coverage: 1059 Selected: 0 ID: Formula

## ToxPrint (729)

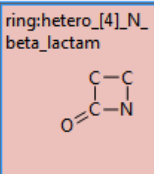
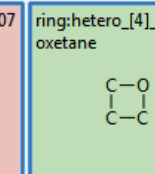
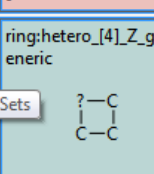
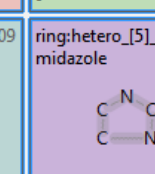
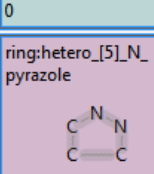
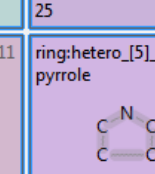
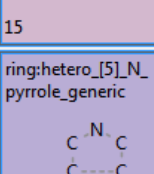
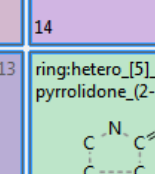
toxprint\_v2.0\_r212.xml

Chemotype Sets

- ToxPrint Chemotypes Version 2.0
  - atom
    - element
      - atom:element\_main\_group
      - atom:element\_metal\_group...
      - atom:element\_metal\_group...
      - atom:element\_metal\_met...
      - atom:element\_metal\_poor...
      - atom:element\_metal\_trans...
      - atom:element\_noble\_gas

Filter Chemotypes According to Chemotype Sets

- C#N
- C(~Z)~C~Q
- C(=O)N
- C(=O)O
- C=N
- C=O
- C=S
- CC(=O)C
- CN
- CNO
- COC
- COH
- CS
- CX
- N(=O)
- N![C]
- N=[N+]=[N-]
- N=C=O
- N=N
- N=O

 ring:hetero_[4]_N_ 607 beta_lactam 0	 ring:hetero_[4]_O_ 608 oxetane 0
 ring:hetero_[4]_Z_g 609 eneric 0	 ring:hetero_[5]_N_i 610 midazole 25
 ring:hetero_[5]_N_ 611 pyrazole 15	 ring:hetero_[5]_N_ 612 pyrrole 14
 ring:hetero_[5]_N_ 613 pyrrole_generic 42	 ring:hetero_[5]_N_ 614 pyrrolidone_(2-) 19

607 / 729

Filter Chemotypes by ID

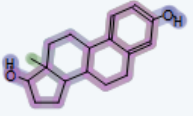
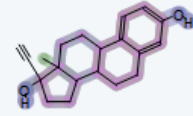
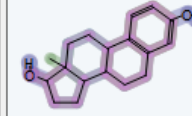
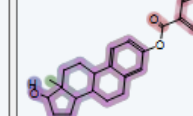
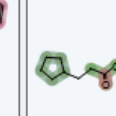
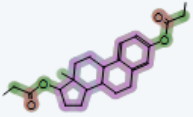
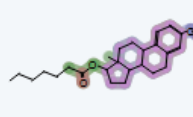
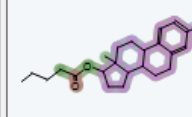
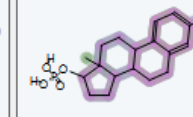
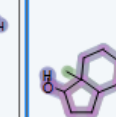
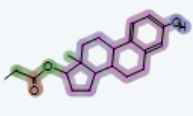
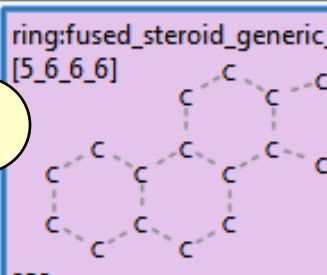
Filter Structures

Chemotypes Loaded: 729 Total Coverage: 467 Selected: 729 ID: Auto

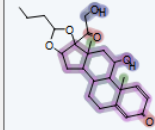
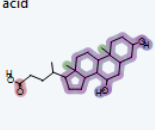
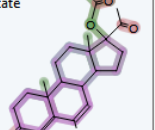
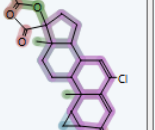
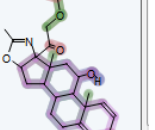
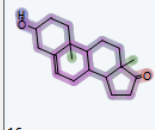
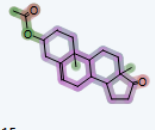
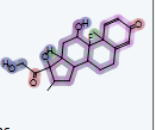
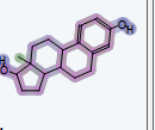
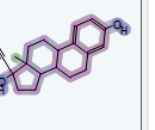
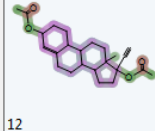
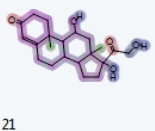
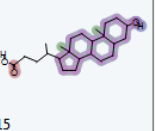
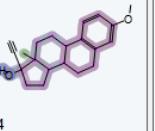
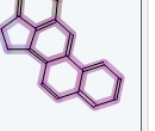
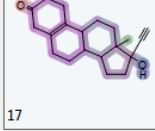
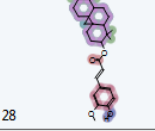
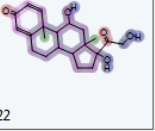
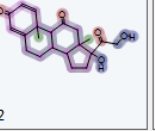
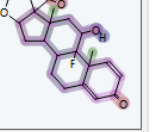
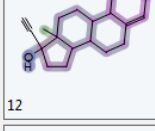
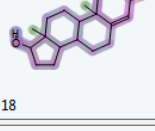
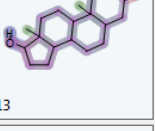
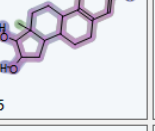
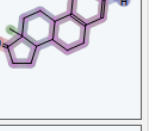
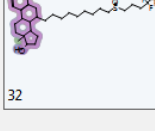
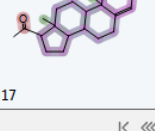
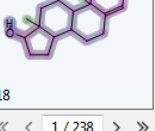
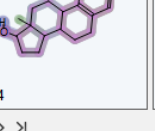
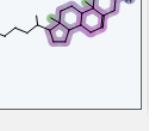


# e.g. Tox21 Estradiol chemotype search

TOX21S\_v4a\_8599\_11Dec2013.sdf

17beta-Estradiol 350 	17alpha-Ethinylestradiol 352 	17alpha-Estradiol 1248 	Estradiol benzoate 1507 	Estradiol cypionate 1507 
Estradiol dipropionate 1509 	Estradiol enanthate 1510 	Estradiol valerate 1511 	Estradiol phosphate 5343 	Estradiol acetate 1507 
beta-Estradiol 17-propionate 7855 	<p>ring:fused_steroid_generic [5_6_6_6]</p>  <p>238</p>			

TOX21S\_v4a\_8599\_11Dec2013.sdf

Budesonide 114 	Chenodeoxycholic acid 153 	Chlormadinone acetate 162 	Cyproterone acetate 226 	Deflazacort 236 
Dehydroepiandrosterone 237 	Dehydroepiandrosterone acetate 238 	Dexamethasone 241 	17beta-Estradiol 350 	17alpha-Ethinylestradiol 352 
Ethinodiol diacetate 374 	Hydrocortisone 437 	Lithocholic acid 470 	Mestranol 488 	3-Methylcholanthrene 516 
Norethynodrel 611 	gamma-Oryzanol 616 	Prednisolone 680 	Prednisone 681 	Triamcinolone acetate 799 
Lynestrenol 858 	Epitestosterone 1234 	5alpha-Dihydrotestosterone 1242 	Estriol 1243 	Estrone 1244 
Fulvestrant 1245 	Progesterone 1246 	Testosterone 1247 	17alpha-Estradiol 1248 	Cholesterol 1257 

1 / 238

Filter Structures by ID type ID Filter Pattern

Filter Chemotypes

Structures Loaded: 8599 Total Coverage: 8454 Selected: 1 Matched: 238 ID: TS\_ChemName

1. Search "estradiol"
2. Identify common chemotype (steroid backbone)
3. Find all chemicals containing chemotype (238/8599 hits)

3

# e.g. ToxCast (1860) "Bisphenol A" chemotype search (25 hits)

The screenshot displays a web-based search interface for ToxCast data. The main area is a grid of 25 chemical structures, each with its name and a numerical ID. The structures are arranged in a 5x5 grid. The first row includes Clorophene (38), Bisphenol A (44), FD&C Green No. 3 (173), 2,2'-Methylenebis(4-methyl-6-tert-butylphenol) (215), and 2,2',6,6'-Tetrachlorobisphenol A (370). The second row includes Dichlorophen (386), 2,2-Bis(4-hydroxyphenyl)-1,1,1-trichloroethane (476), 2-Hydroxy-4-methoxybenzophenone (494), 2,4-Dihydroxybenzophenone (495), and 4,4'-Methylenebis(2,6-di-tert-butylphenol) (498). The third row includes Diphenolic acid (502), Phenolphthalin (504), Bisphenol B (505), Bisphenol F (506), and 4-Cumylphenol (525). The fourth row includes 3,3',5,5'-Tetrabromobisphenol A (835), Octabenzon (981), 2,4-Bis(1-methyl-1-phenylethyl)phenol (1063), 4-Hydroxytamoxifen (1353), and 4,4',4''-Ethane-1,1,1-triyltriphenol (1369). The fifth row includes a partially visible structure (1460), Sulisobenzon (1583), and 2-Hydroxybenzophenone (1801). On the right side, there is a sidebar with search results. The first result is 'bond:COH\_alcohol\_aromatic 123' with a structure showing a hydroxyl group on a benzene ring. The second result is 'chain:aromaticAlkane\_Ph-C1-Ph 479' with a structure showing two benzene rings connected by a methylene group. At the bottom right, there is a search filter: 'Filter Chemotypes by ID' with a text input field containing 'type ID...'. A yellow text box is overlaid on the bottom left of the image.

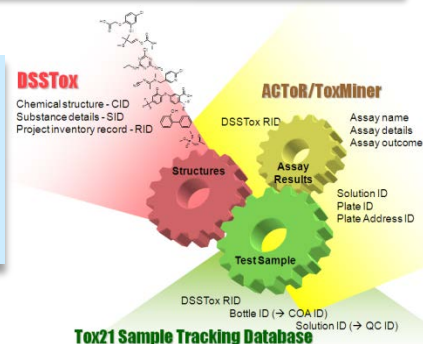
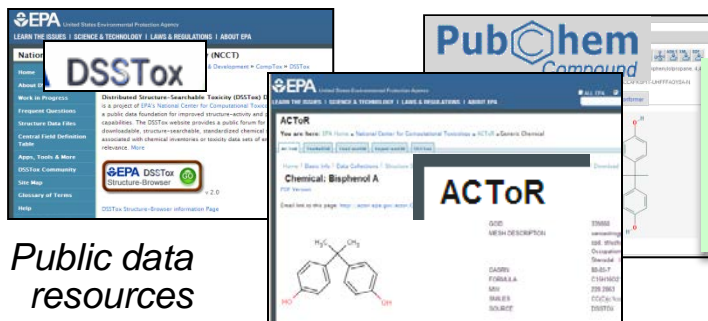
Can export list of chemotypes for selected chemicals

Can export structures containing chemotypes

→ Use in iCSS Dashboard to explore ToxCast HTS results

# Chemistry: What's needed

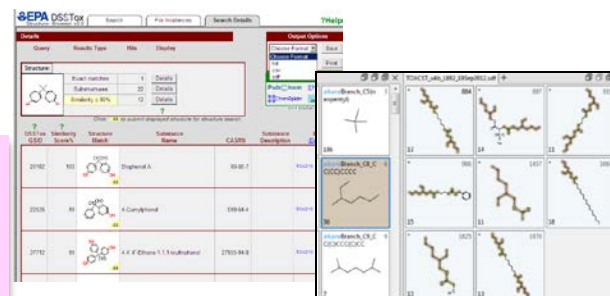
- Accurate chemical annotations of testing libraries (e.g., ToxCast & Tox21), transparency, & reporting of error sources
- Cheminformatics foundation to enable structure modeling

**Public data resources**

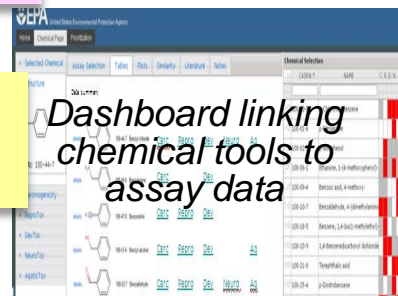
Public data release: ability of chemists & non-chemists (biologists, statisticians) to access & utilize chemical information

Structure-similarity searching



Use all available data (HTS+chemistry) to form hypotheses, guide & inform analog selection, and improve prediction models

Incorporate chemical information into usable tools for chemical prioritization & safety assessments



**Dashboard linking chemical tools to assay data**

Informed feature-grouping & highlighting

# Questions?