EPA United States Environmental Protection Agency

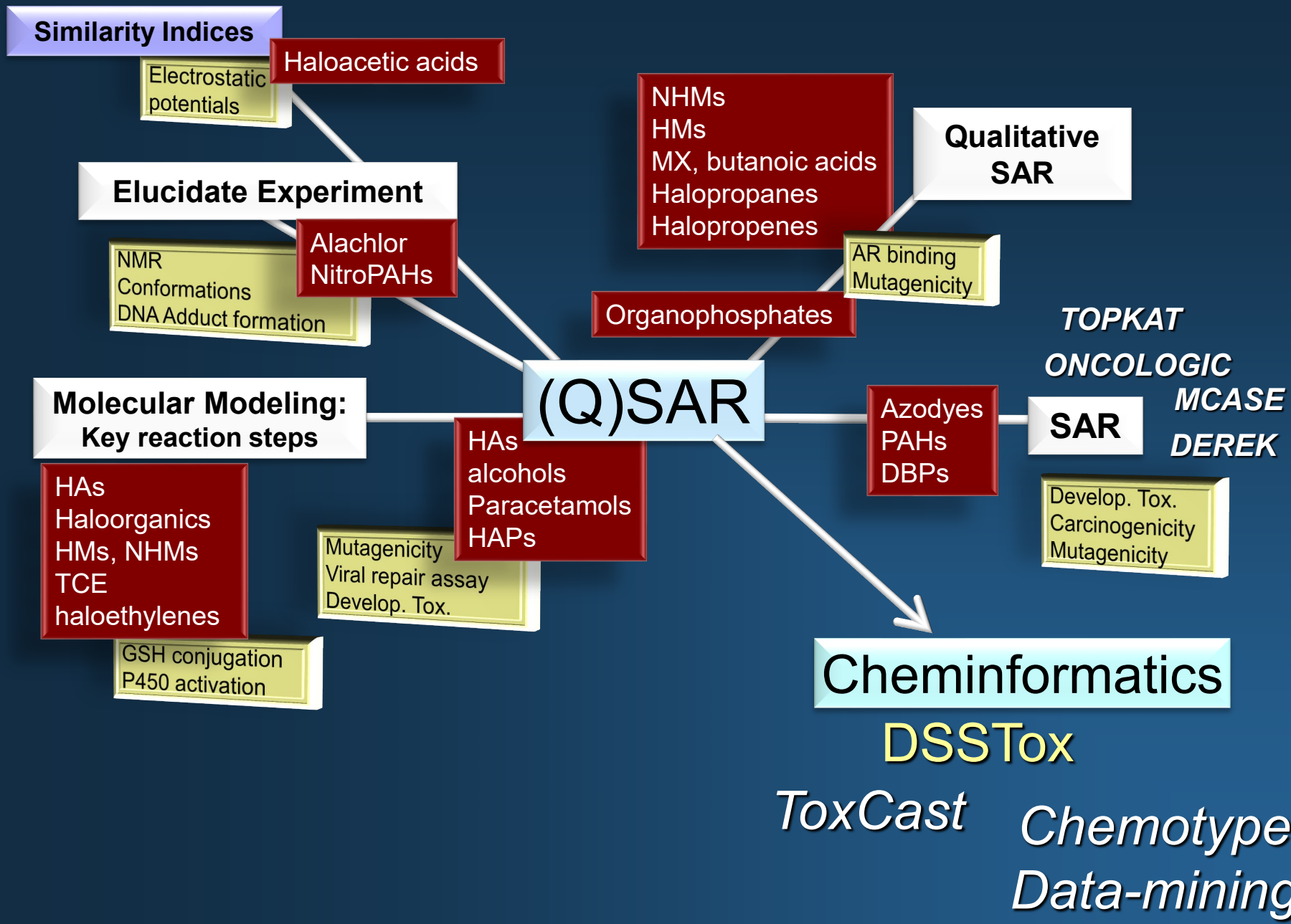
SUNY Buffalo, Chemistry Dept. Seminar, March 7th, 2019

EPA's Computational Toxicology Program: Innovation Powered by Chemistry

Ann Richard
National Center for Computational Toxicology
Office of Research & Development
richard.ann@epa.gov

Office of Research and Development
National Center for Computational Toxicology

Past & present research at EPA:

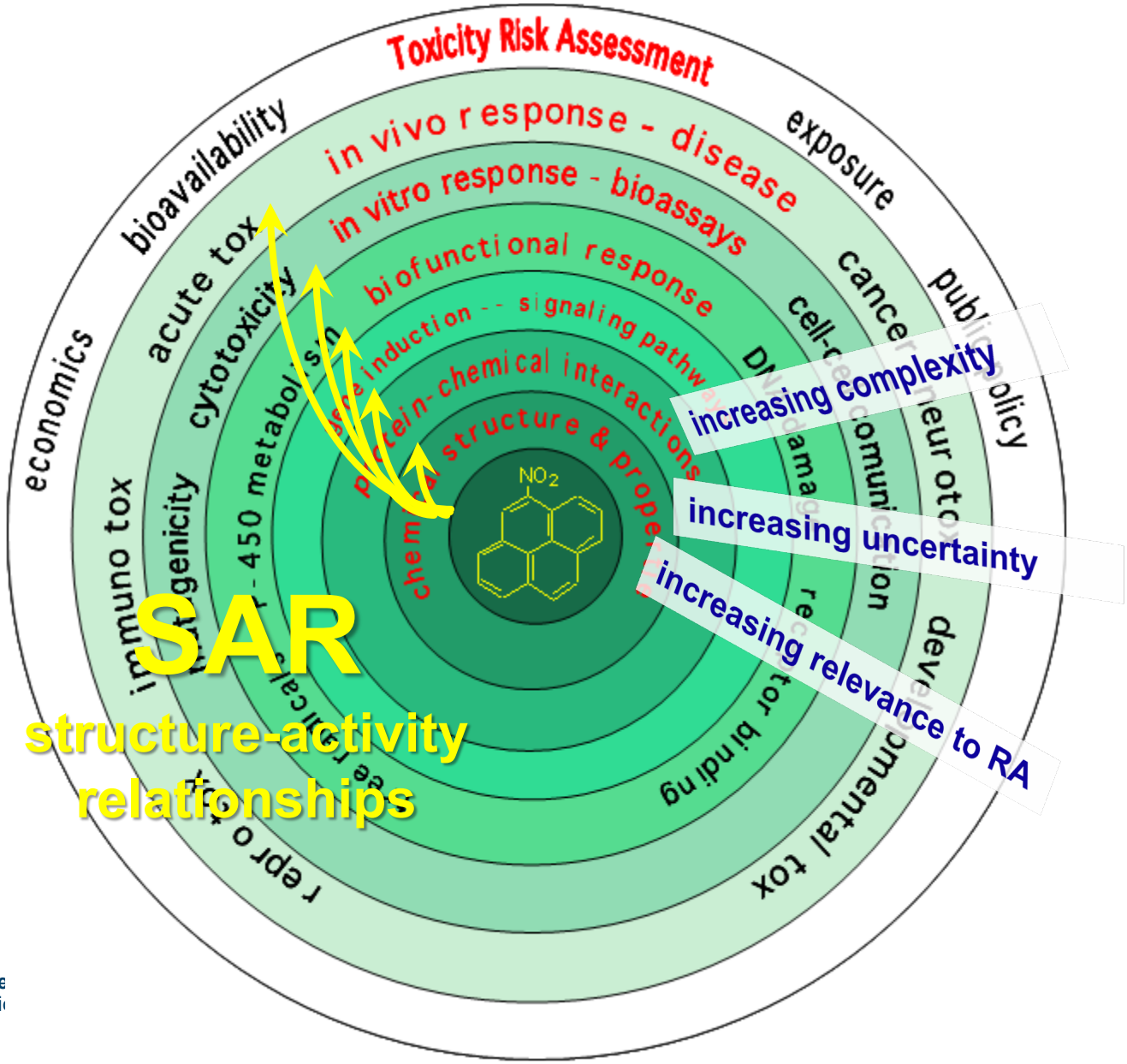


#1:

"Asking the right questions takes as much skill as giving the right answers."

- Robert Half

Toxicity Prediction Problem



United States Environmental Protection Agency

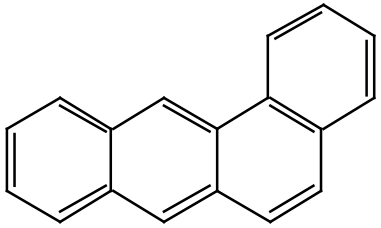
Structure-**A**ctivity **R**elationship (**SAR**) postulate:

similar molecules have similar activities

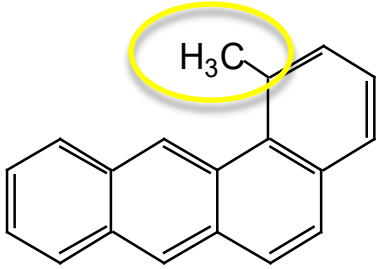
(except when they don't!)

Structure-Activity Relationships (SAR)

Activity = f (Structure)



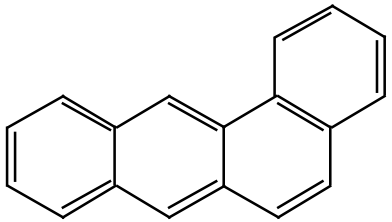
active



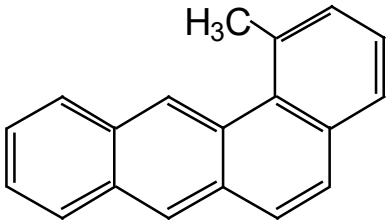
inactive

Methyl group leads to loss of carcinogenic activity

SAR Generalization



active



inactive

Statistical association

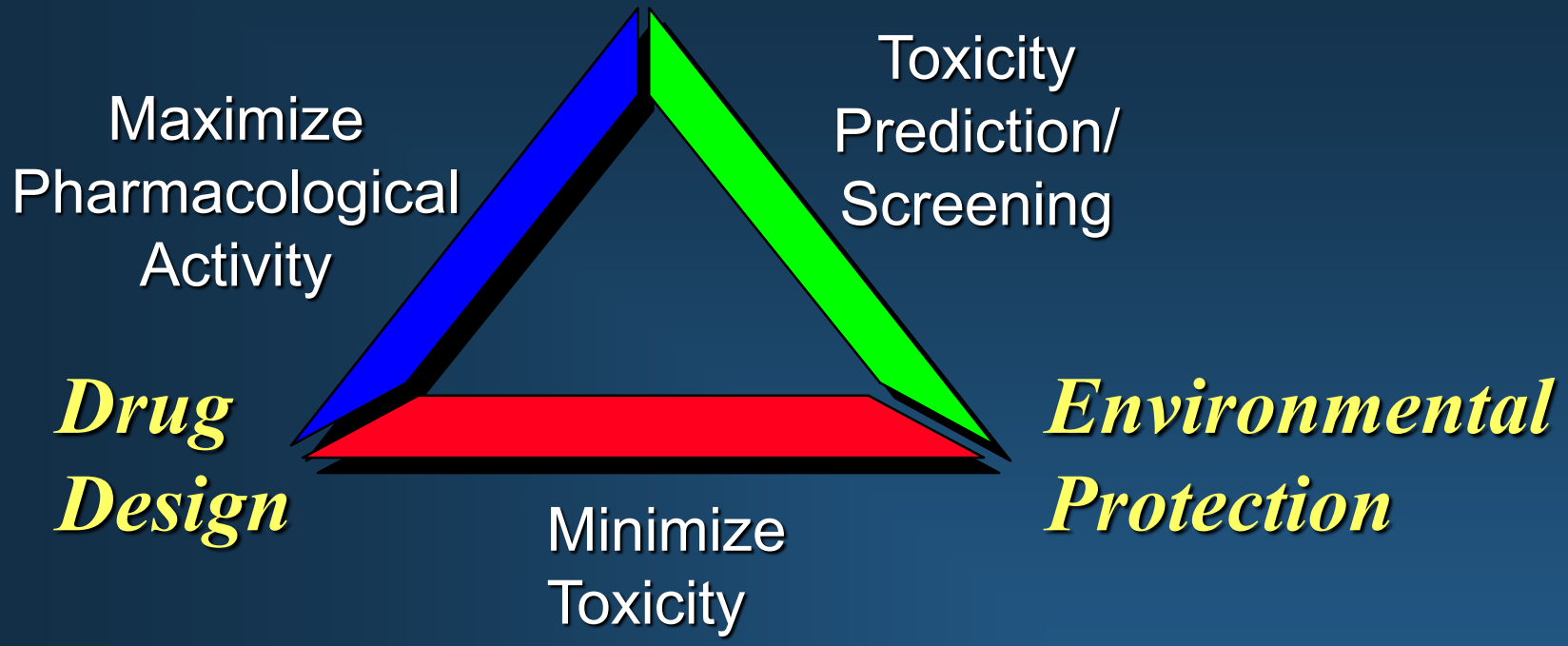
Mechanistic hypothesis

Class → PAHs

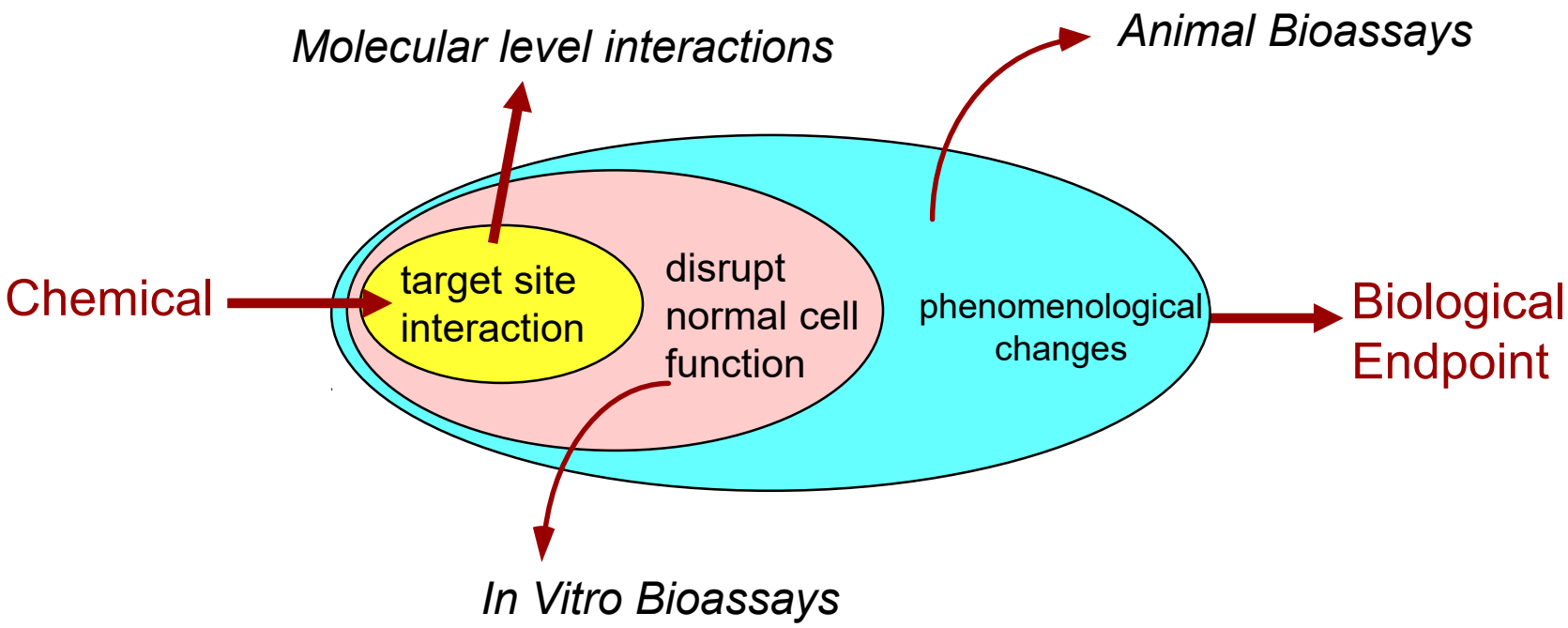
activating feature → bay region

modulating feature → steric hindrance

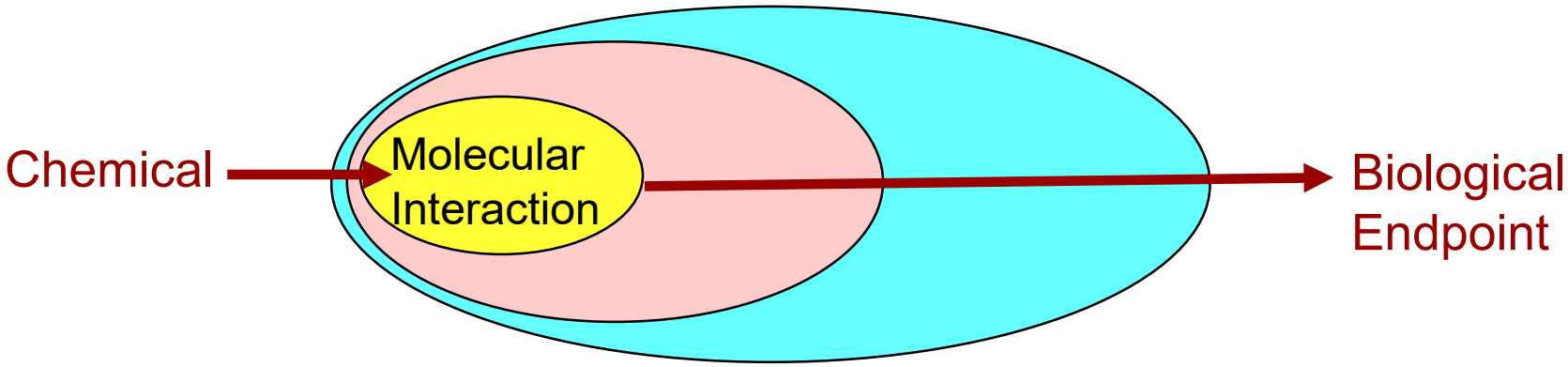
SAR Application



Mechanisms of Toxicity



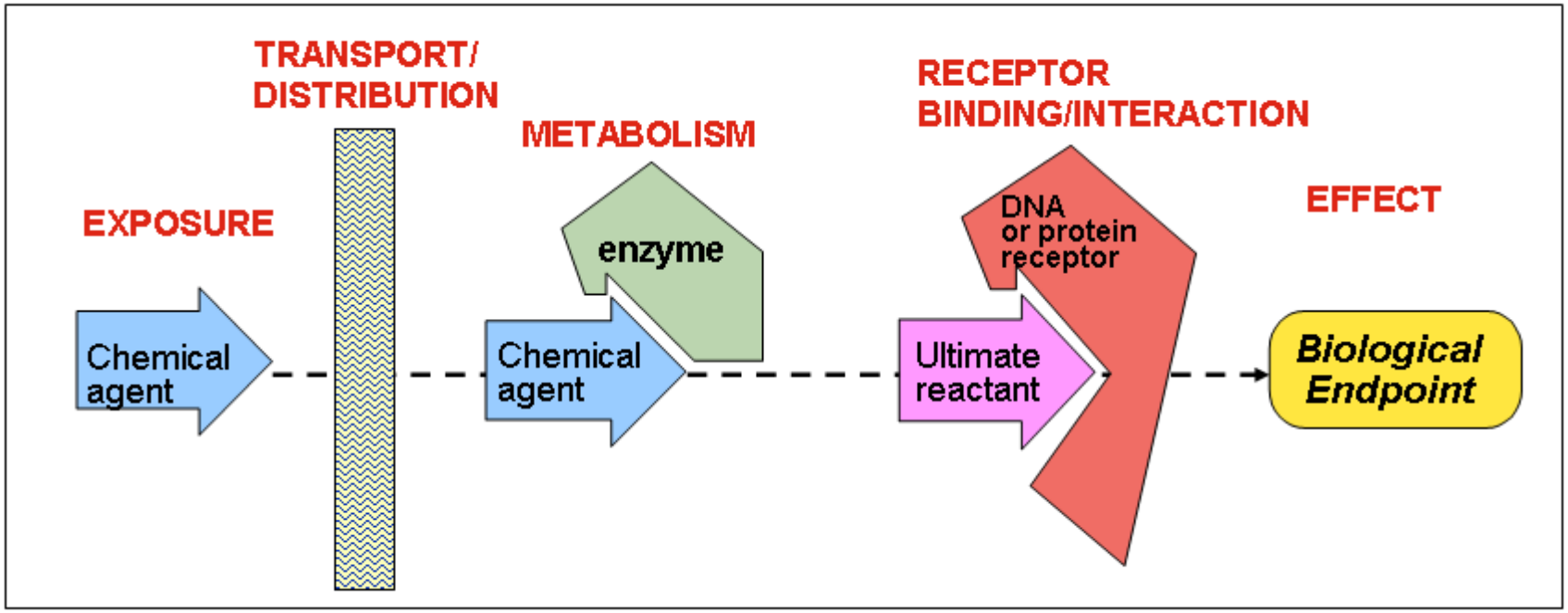
Structure-Activity Relationships



Similar chemicals
Relative properties
Common mechanism of action

Typical (Q)SAR Paradigm

$$\text{Activity} = \text{Constant} * \text{Prob}_{site} * \text{Prob}_{rxn}$$

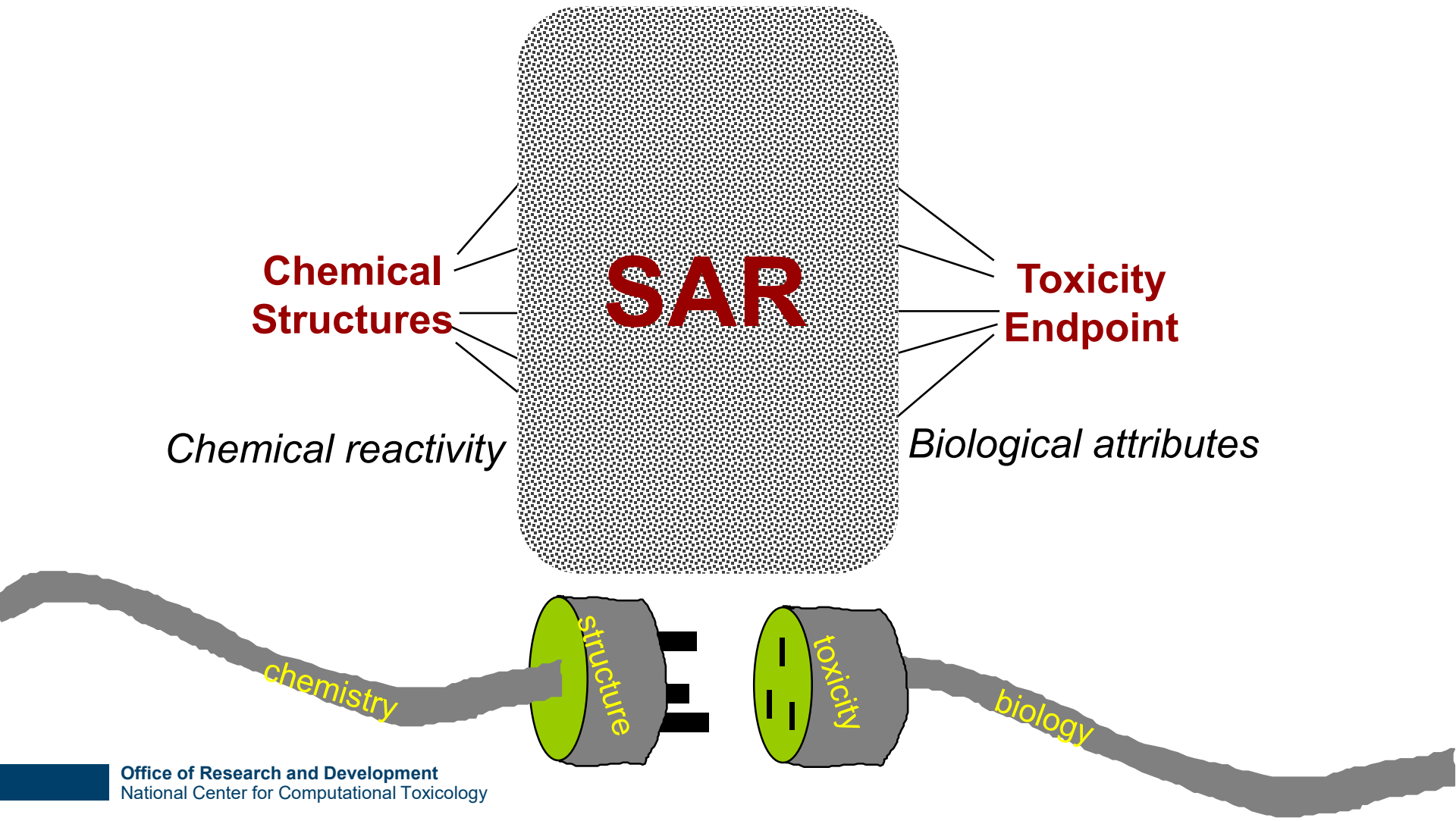


vapor pressure
Solubility
Henry's const

log P(octanol/water)
acidity

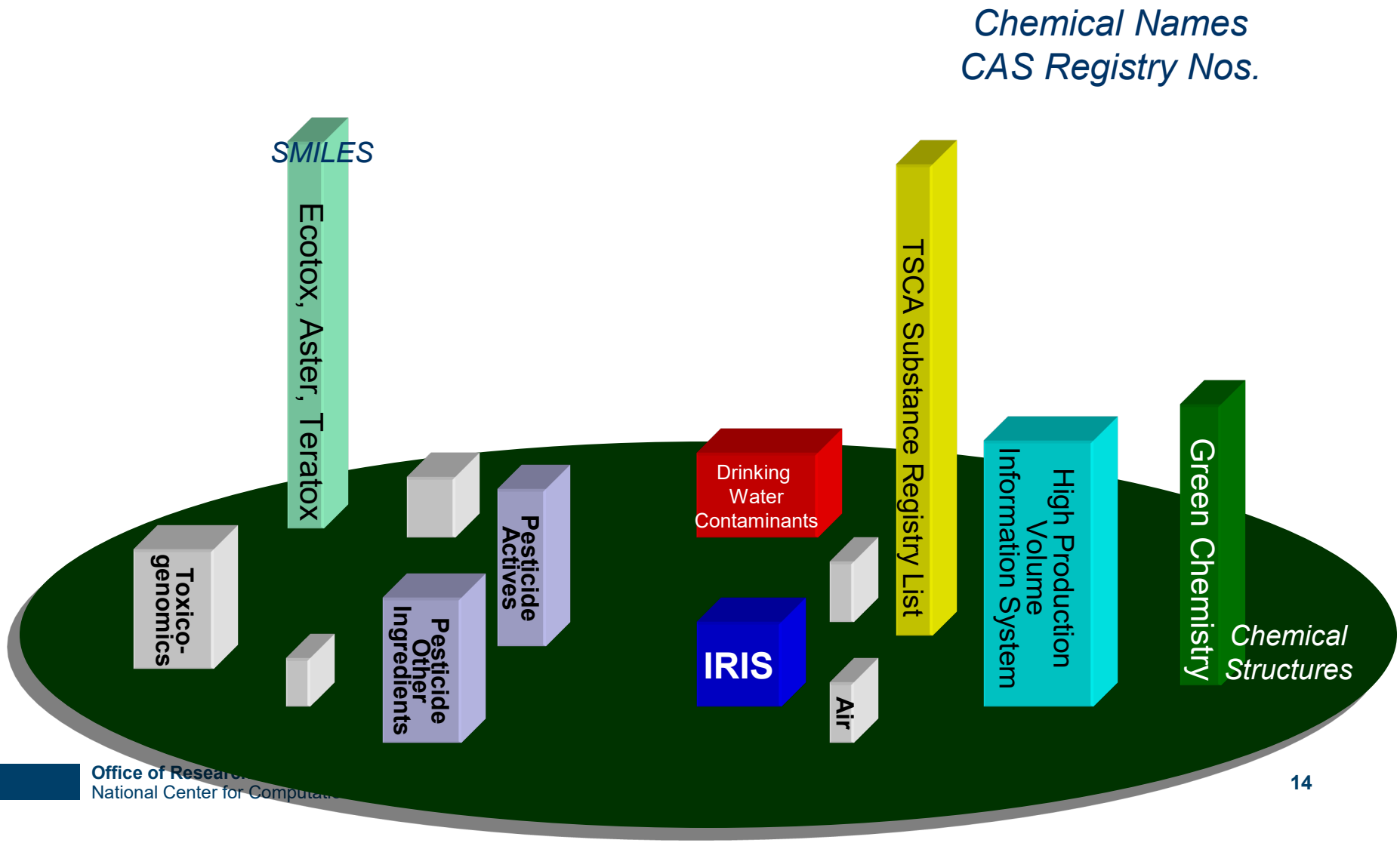
electronic/ steric
3D properties
reactivity indicators
interaction energies

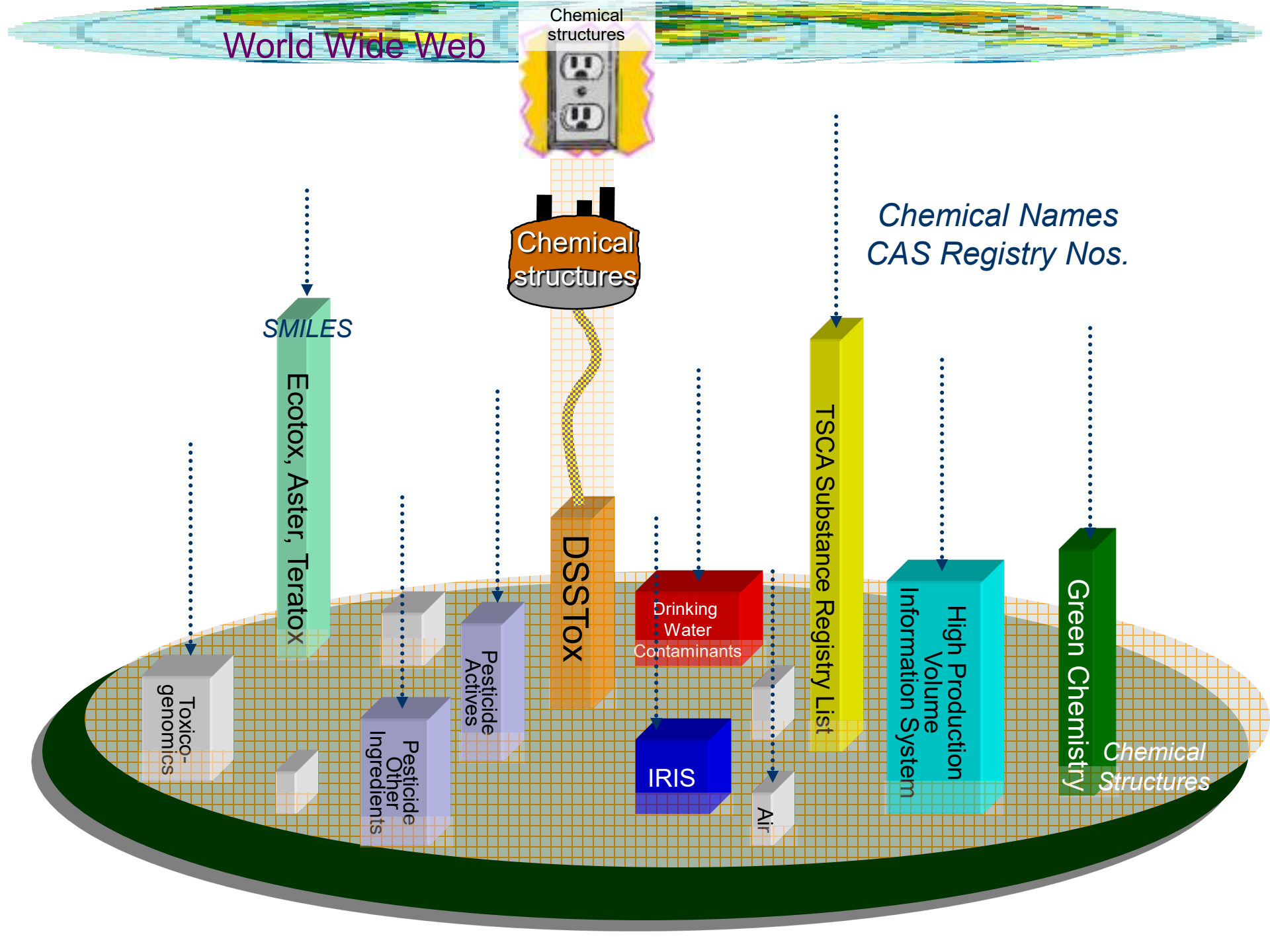
Global vs. Local SAR models



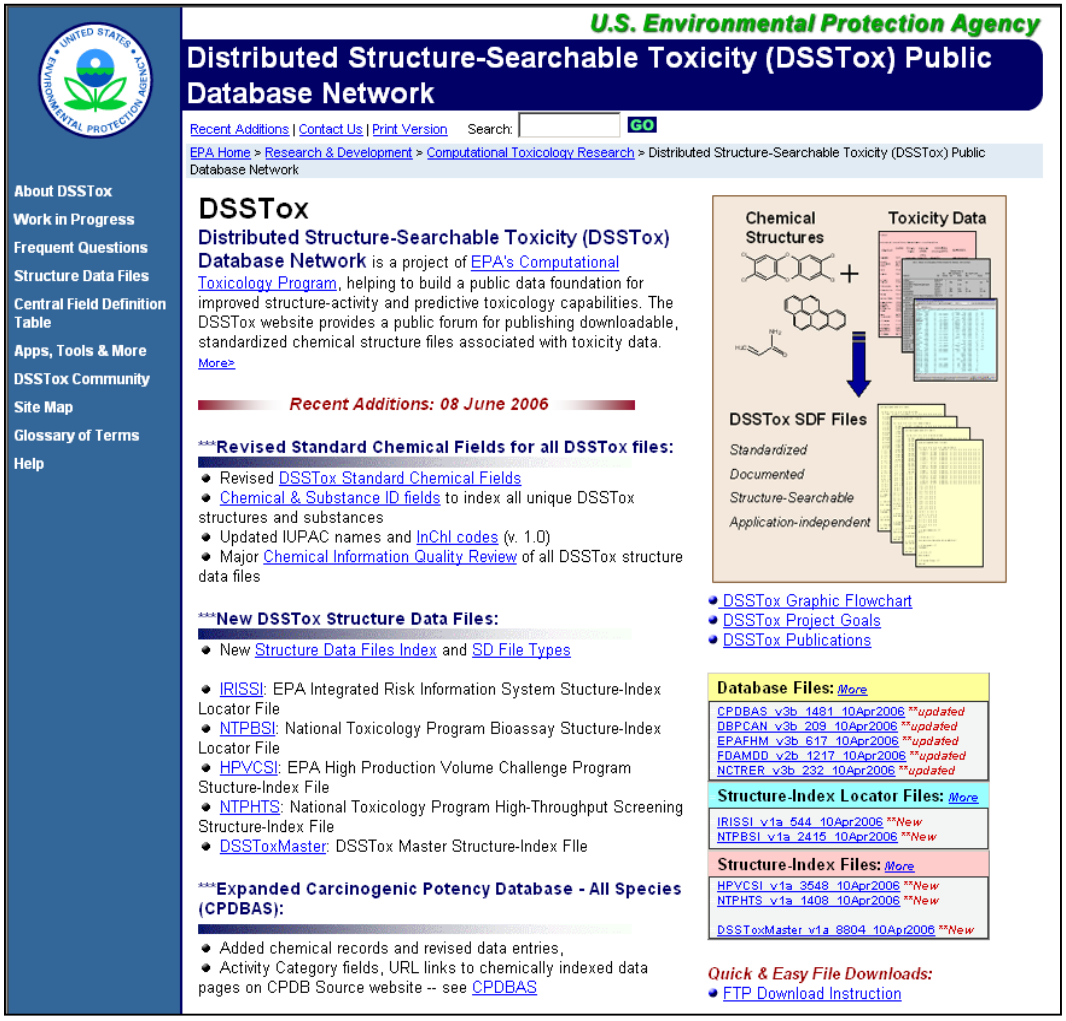
#2:
Biology is complicated and predicting chemical toxicity is hard

EPA's data islands … circa 2000




World Wide Web
Chemical structures
Chemical structures
Chemical Names
CAS Registry Nos.
SMILES
Ecotox, Aster, Teratox
DSSTox
TSCA Substance Registry List
High Production Volume Information System
Green Chemistry
Chemical Structures
Drinking Water Contaminants
IRIS
Air
Toxico-genomics
Pesticide Actives
Pesticide Other Ingredients

EPA's DSSTox Public Website – Launched 2004



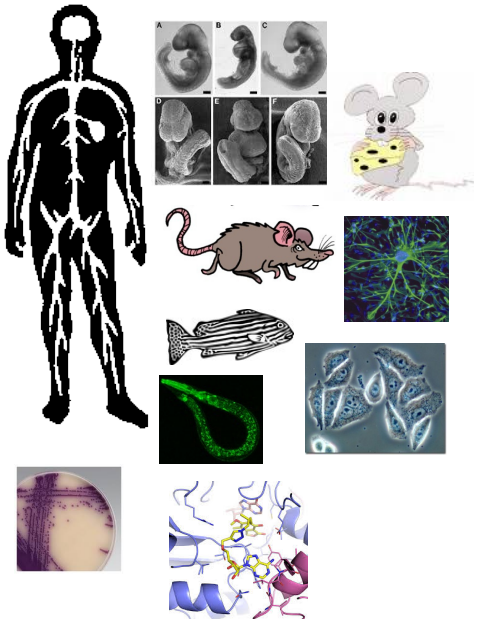
- Focus on environmental chemicals & EPA lists
- Toxicity datasets for building SAR models
- High quality, manually curated data-structure associations
- Downloadable structure files

Distributed structure-searchable toxicity (DSSTox) database network: A proposal. Richard, A.M., Williams, C.R. Mut. Res., 499:27-52, ***2002****.*

Improving structure-linked access to publicly available chemical toxicity information. Richard, A.M., Williams, C.R., Cariello, N. Curr. Opinion Drug Devel Discov., 5:136-143, ***2002****.*

DSSTox Website launch: Improving public access to databases for building structure-toxicity prediction models. Richard, A.M. Preclinica, 2:103-108, ***2004****.*

Errors in Data-Structure Linkages



List

$Data_1$ → CAS_1 ↔ $Name_1$ ↔ $Structure_1$

$Data_2$ → CAS_2 ↔ $Name_2$ ↔ $Structure_1$

$Data_3$ → CAS_3 ↔ $Name_3$ ↔ $Structure_3$

$Data_4$ → CAS_4 ↔ $Name_4$ ↔ $Structure_4$

List

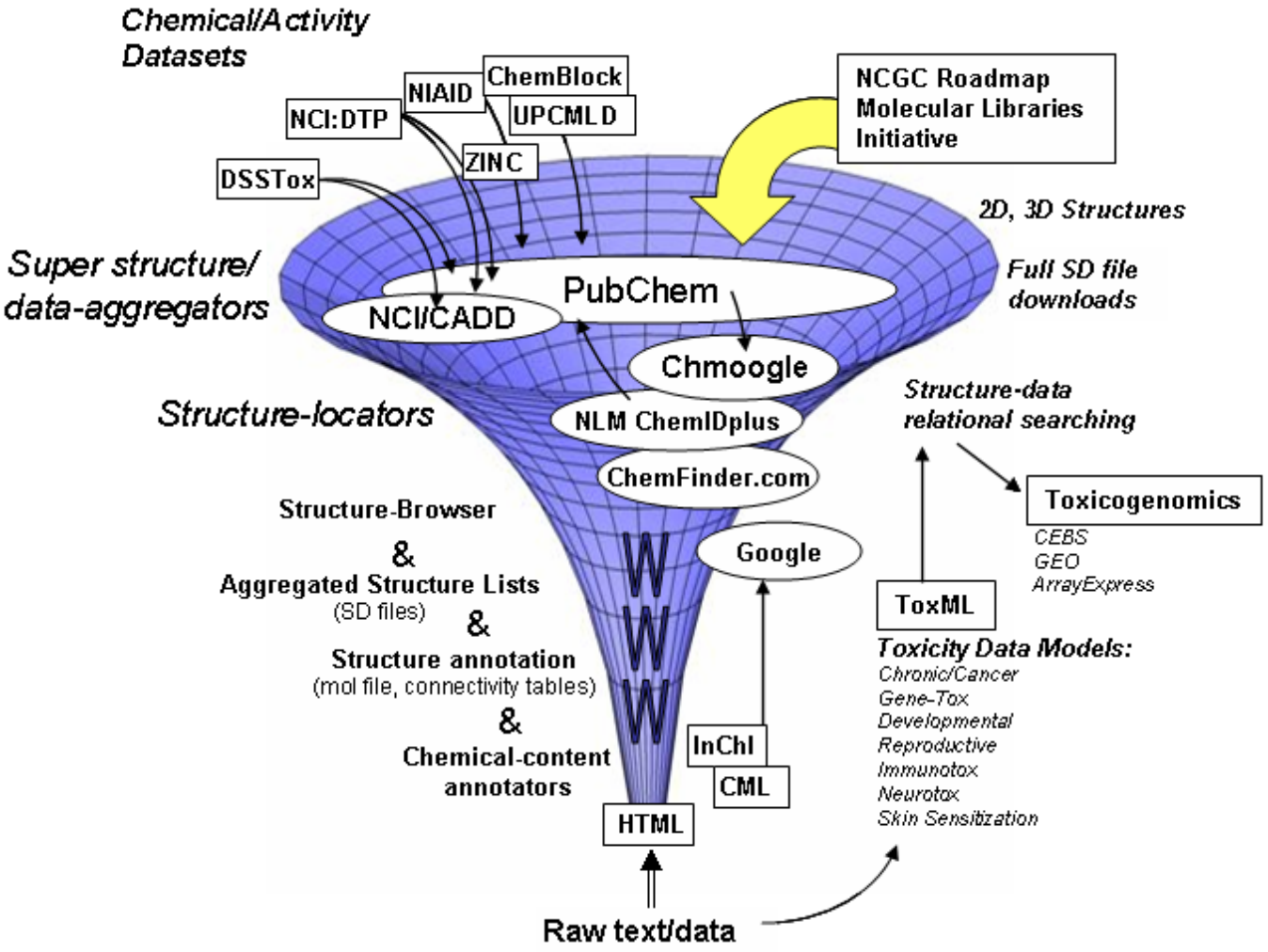
$Data_1$? CAS_1 ? $Name_1$? $Structure_1$

$Data_2$? CAS_2 ? $Name_2$? $Structure_1$

$Data_3$? CAS_3 ? $Name_3$? $Structure_3$

$Data_4$? CAS_4 ? $Name_4$? $Structure_4$

Chemical Structure Searching of Biological Information on the Internet



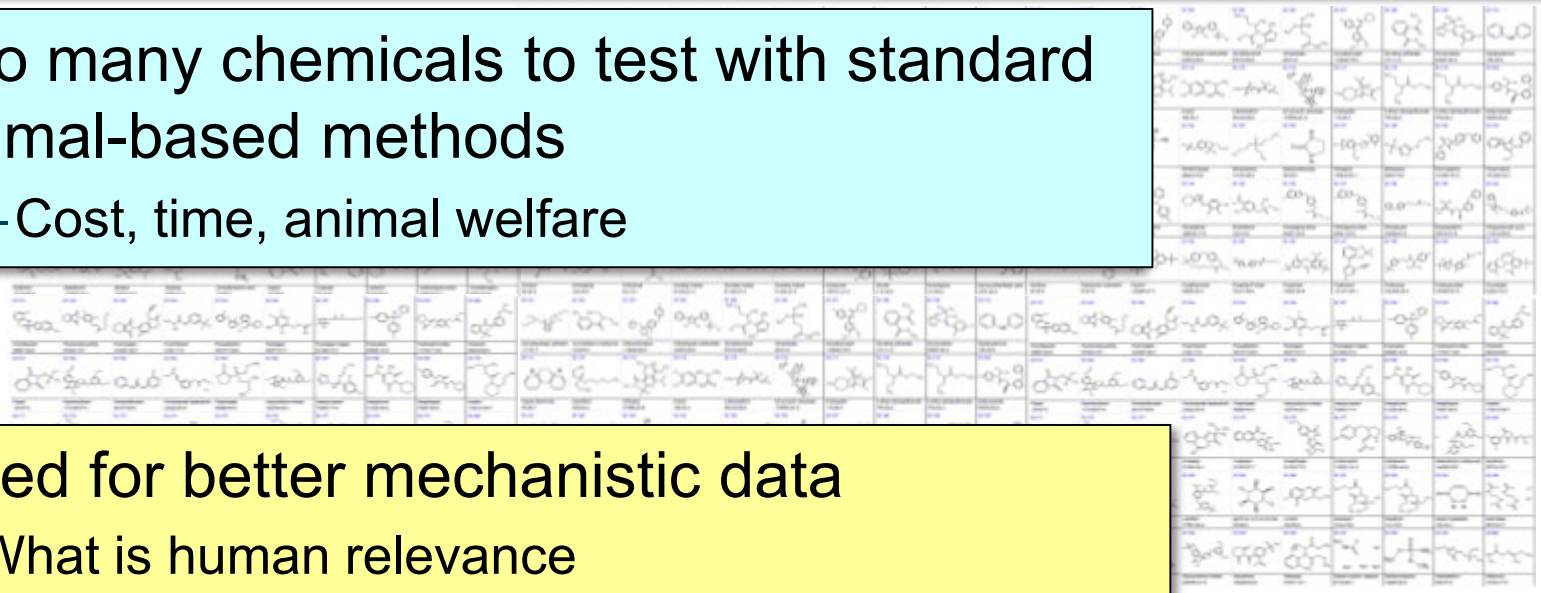
Richard A.M., Gold, L.S., Nicklaus, M.C. (2006) Chemical structure indexing of toxicity data on the internet: Moving towards a flat world. Curr. Opinion Drug Disc. Develop., 9(3): 314-325.

#3:
Importance of data quality

Big Problems in Toxicology

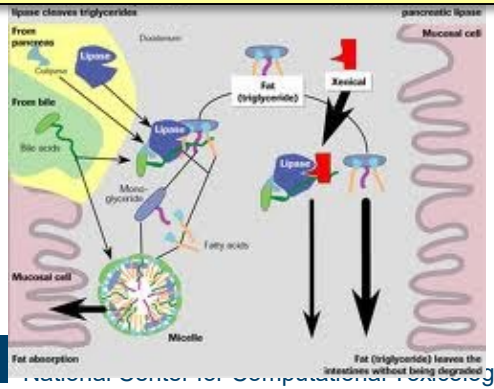
Too many chemicals to test with standard animal-based methods

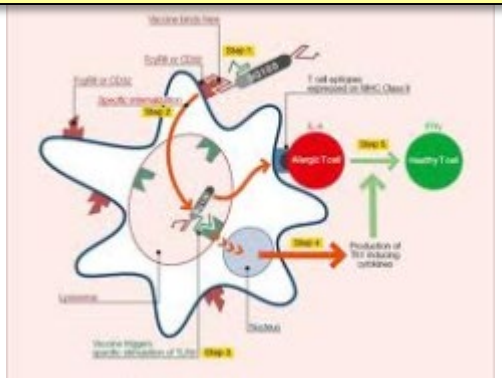
- Cost, time, animal welfare

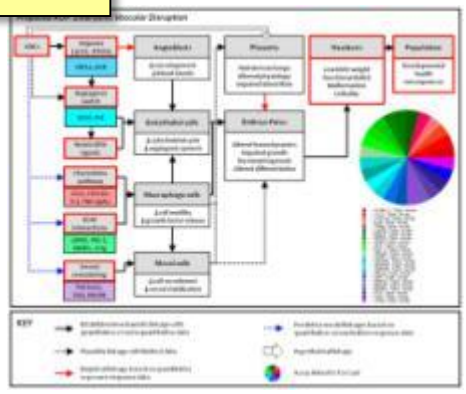


Need for better mechanistic data

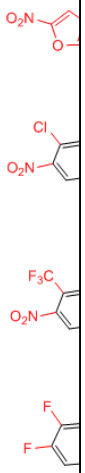
- What is human relevance
- What is the Mode of Action (MOA)?
- What is the Adverse Outcome Pathway (AOP)?







Future of Chemical Toxicity Testing



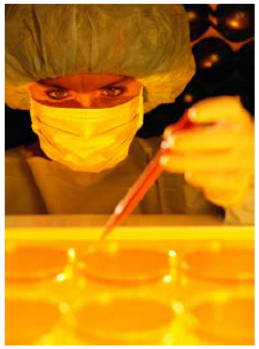
THE NATIONAL REPORT IN BRIEF

July 2007

Toxicity Testing in the 21st Century: A Vision and a Strategy

Advances in molecular biology, biotechnology, and other fields are paving the way for major improvements in how scientists evaluate the health risks posed by potentially toxic chemicals found at low levels in the environment. These advances would make toxicity testing quicker, less expensive, and more directly relevant to human exposures. They could also reduce the need for animal testing by substituting more laboratory tests based on human cells. This National Research Council report creates a far-reaching vision for the future of toxicity testing.

Toxicity tests on laboratory animals are conducted to evaluate chemicals—including medicines, food additives, and industrial, consumer, and agricultural chemicals—for their potential to cause cancer, birth defects, and other adverse health effects. Information from toxicity testing serves as an important part of the basis for public health and regulatory decisions concerning toxic chemicals. Current test methods were developed incrementally over the past 50 to 60 years and are conducted using laboratory animals, such as rats and mice. Using the results of animal tests to predict human health effects involves a number of assumptions and extrapolations that remain controversial. Test animals are often exposed to higher doses than would be expected for typical human exposures, requiring assumptions about effects at lower doses or exposures. Test animals are typically observed for overt signs of adverse health effects, which provide little information about biological changes leading to such health effects. Often controversial uncertainty factors must be applied to account for differences between test animals and humans. Finally, use of animals in testing is expensive and time consuming, and it sometimes raises ethical issues.



Today, toxicological evaluation of chemicals is poised to take advantage of the on-going revolution in biology and biotechnology. This revolution is making it increasingly possible to study the effects of chemicals using cells, cellular components, and tissues—preferably of human origin—rather than whole animals. These powerful new approaches should help to address a number of challenges facing the

National Academ

POLICYFORUM

TOXICOLOGY

Transforming Environmental Health Protection

We propose a shift from primarily in vivo animal studies to in vitro assays, in vivo assays with lower organisms, and computational modeling for toxicity assessments.

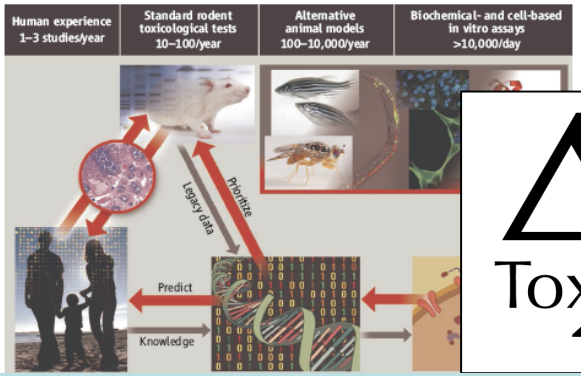
Francis S. Collins,[1*†] George M. Gray,[2*] John R. Bucher[3*]

In 2005, the U.S. Environmental Protection Agency (EPA), with support from the U.S. National Toxicology Program (NTP), funded a project at the National Research Council (NRC) to develop a long-range vision for toxicity testing and a strategic plan for implementing that vision. Both agencies wanted future toxicity testing and assessment paradigms to meet evolving regulatory needs. Challenges include the large numbers of substances that need to be tested and how to incorporate recent advances in molecular toxicology, computational sciences, and information technology; to rely increasingly on human as opposed to animal data; and to offer increased efficiency in design and costs (*1*–*5*). In response, the NRC Committee on Toxicity Testing and Assessment of Environmental Agents produced two reports that reviewed current toxicity testing, identified key issues, and developed a vision and implementation strategy to create a major shift in the assessment of chemical hazard and risk (*6*, *7*). Although the NRC reports have laid out a solid theoretical rationale, comprehensive and rigorously gathered data (and comparisons with historical animal data) will determine whether the hypothesized improvements will be realized in practice. For this purpose, NTP, EPA, and the National Institutes of Health Chemical Genomics Center (NCGC) (organizations with expertise in experimental toxicology, computational toxicology, and high-throughput technologies, respectively) have established a collaborative research program.

EPA, NCGC, and NTP Joint Activities

In 2004, the NTP released its vision and roadmap for the 21st century (*1*), which established initiatives to integrate high-throughput screening (HTS) and other automated screening assays into its testing program. In 2005, the EPA established the National Center for Computational Toxicology (NCCT). Through these initiatives, NTP and EPA, with the NCGC, are promoting the evolution of toxicology from a predominantly observational science at the level of disease-specific models in vivo to a predominantly predictive science focused on broad inclusion of target-specific, mechanism-based, biological observations in vitro (*1*, *4*) (see figure, below).

Toxicity pathways. In vitro and in vivo tools are being used to identify cellular responses after chemical exposure expected to result in adverse health effects (*7*). HTS methods are a primary means of discovery for drug development, and screening of >100,000 compounds per day is routine (*8*). However, drug-discovery HTS methods traditionally test compounds at one concentration, usually between 2 and 10 μM, and tolerate high false-negative rates. In contrast, in the EPA, NCGC, and NTP combined effort, all compounds are tested at as many as 15 concentrations, generally ranging from ~5 nM to ~100 μM, to generate a concentration-response curve (*9*). This approach is highly reproducible, produces significantly lower false-positive and false-negative rates than the traditional HTS methods (*9*), and facilitates multiassay comparisons. Finally, an informatics platform has been built to compare results among HTS screens; this is being expanded to allow comparisons with historical toxicologic NTP and EPA data (http://ncgc.nih.gov/pub/openhts). HTS data collected by EPA and NTP, as well as by the NCGC and other Molecular Libraries Initiative centers (http://mli.nih.gov/), are being made publicly available through Web-based databases [e.g., PubChem (http://pubchem.ncbi.nlm.nih.gov)]. In addition,



Transforming toxicology. The studies we propose will test whether high-throughput and computational toxicology approaches can yield data predictive of results from animal toxicity studies, will allow prioritization of chemicals for further testing, and can assist in prediction of risk to humans.

[1]Director, National Human Genome Research Institute (NHGRI), National Institutes of Health, Bethesda, MD 20892; [2]Assistant Administrator for the Office of Research and Development, U.S. Environmental Protection Agency, policies of their respective agencies.

†Author for correspondence. E-mail: francisc@mail.nih.gov

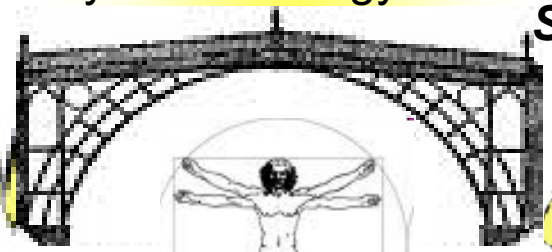
Tox21

EPAs Contribution: The ToxCast Research Program

Drug Discovery

Chemical libraries
HTS screening
Bioprofiling
Chemical probes
Molecular profiling
Feature optimization
Cheminformatics
Chemogenomics
Systems biology

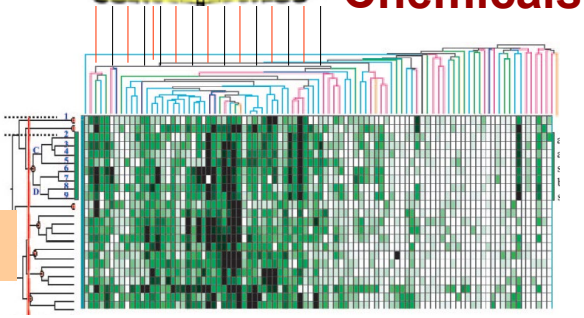
Chemical Candidates, Scaffolds, Virtual libraries



Toxicity Screening

Drugs, Industrial, & Environmental Chemicals

Therapeutic Endpoint



Polypharmacology | Polyfunctional

Low dose | High dose

Structure modification | Green chemistry

High specificity
High affinity

Chronic, Acute, Devel,
Repro, Immuno, Neuro…tox

Low specificity
Low affinity

High-Throughput Screening Assays

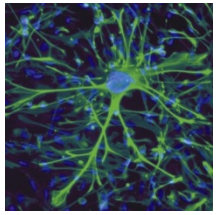
batch testing of chemicals for pharmacological/toxicological endpoints using automated liquid handling, detectors, and data acquisition

LTS MTS HTS uHTS

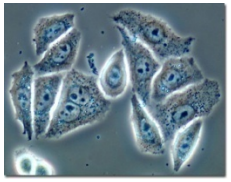
10s-100s/yr

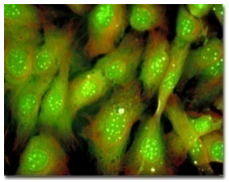
Gene-expression

10s-100s/day

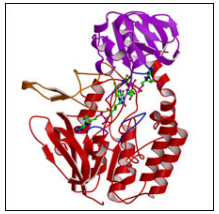


1000s/day





10,000s-100,000s/day



Human Relevance/ Cost/Complexity

Throughput/ Simplicity

ToxCast Assays

United States Environmental Protection Agency

Biochemical Assays

- Protein families
  - GPCR
  - NR
  - Kinase
  - Phosphatase
  - Protease
  - Other enzyme
  - Ion channel
  - Transporter
- Assay formats
  - Radioligand binding
  - Enzyme activity
  - Co-activator recruitment

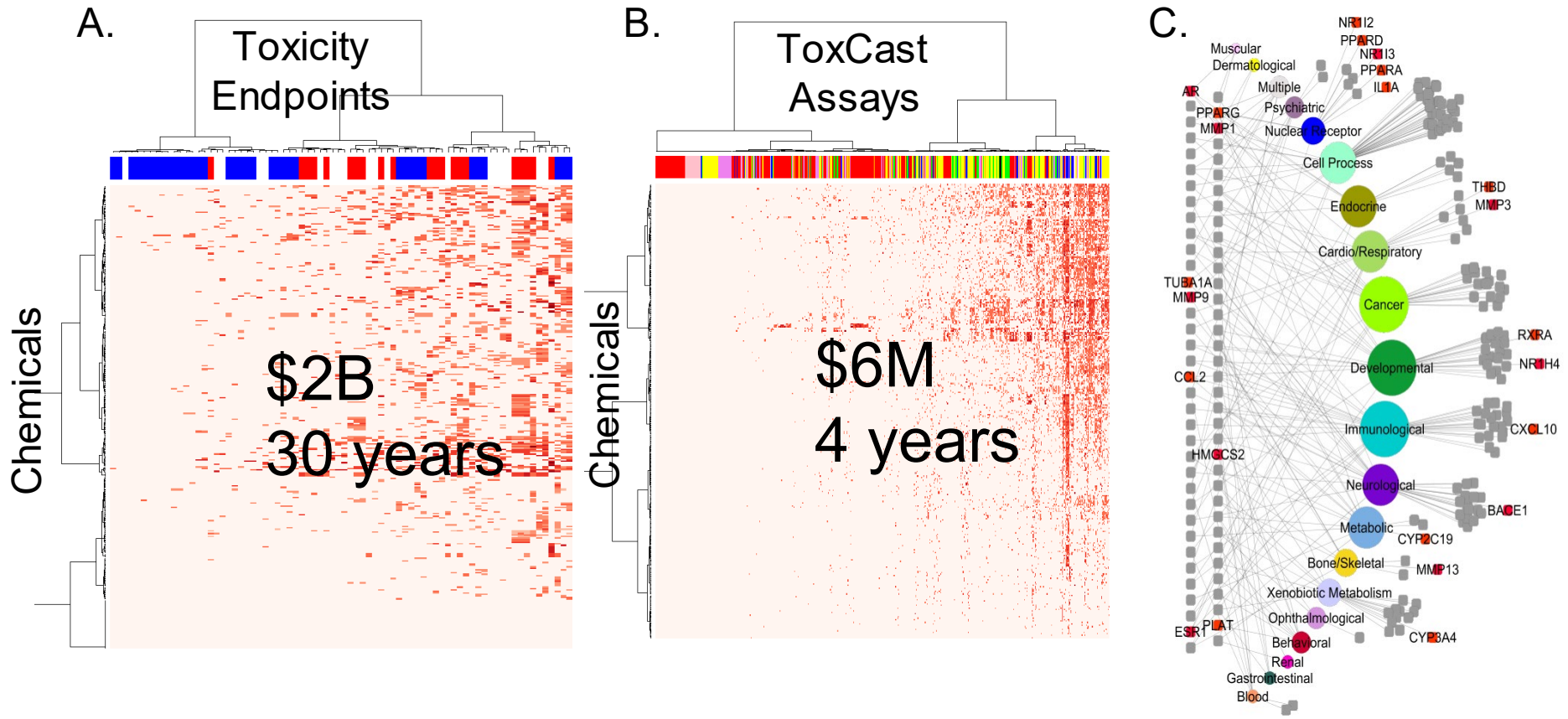
~1200 Total Endpoints

Primarily Human / Rodent
Exception: Zebrafish development (S. Padilla)

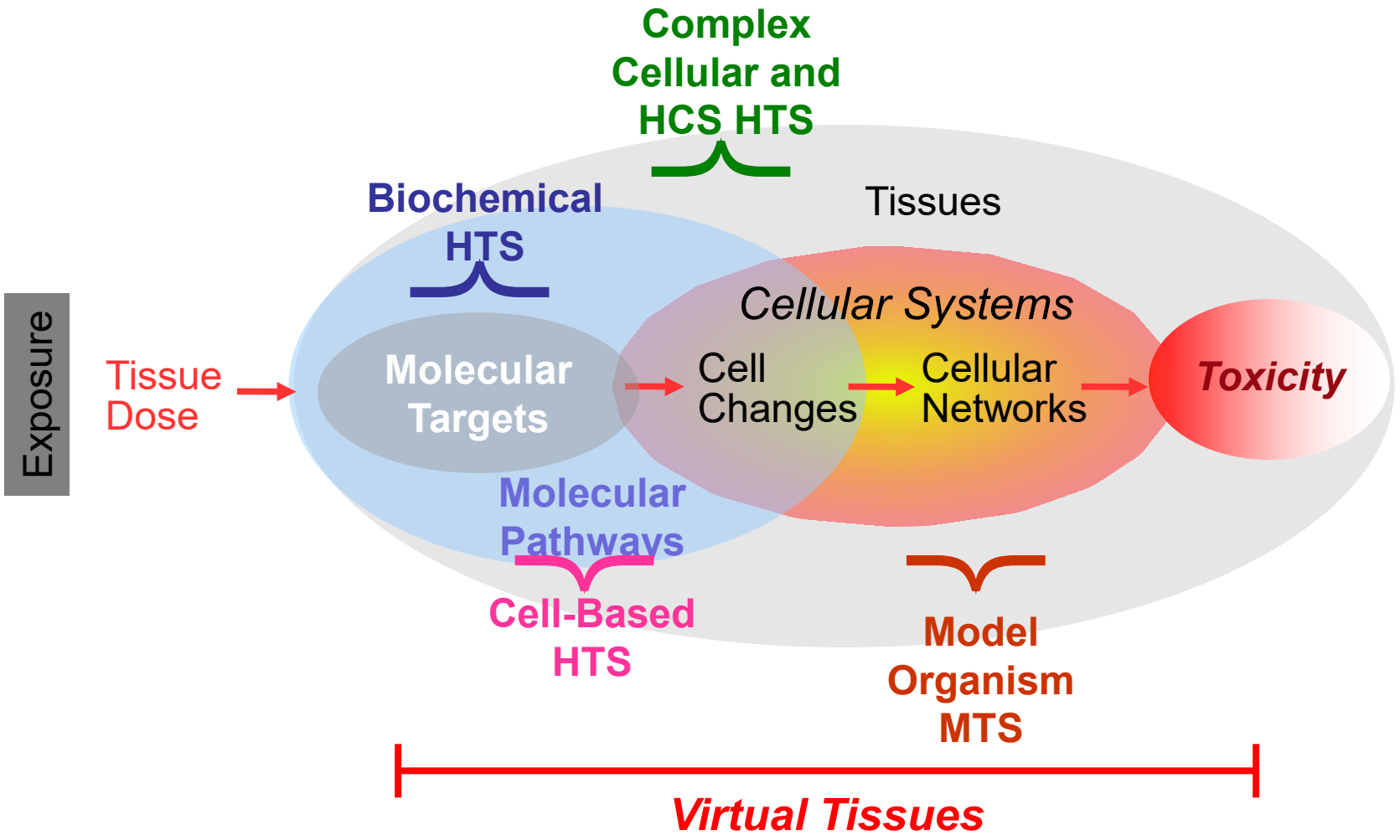
Cellular Assays

- Cell lines
  - HepG2 human hepatoblastoma
  - A549 human lung carcinoma
  - HEK 293 human embryonic kidney
- Primary cells
  - Human endothelial cells
  - Human monocytes
  - Human keratinocytes
  - Human fibroblasts
  - Human proximal tubule kidney cells
  - Human small airway epithelial cells
  - Rat hepatocytes
  - Mouse embryonic stem cells (Sid Hunter)
- Biotransformation competent cells
  - Primary rat hepatocytes
  - Primary human hepatocytes
- Assay formats
  - Cytotoxicity
  - Reporter gene
  - Gene expression
  - Biomarker production
  - High-content imaging for cellular phenotype

ToxRefDB ⟵ ToxCast ⟶ Human Disease

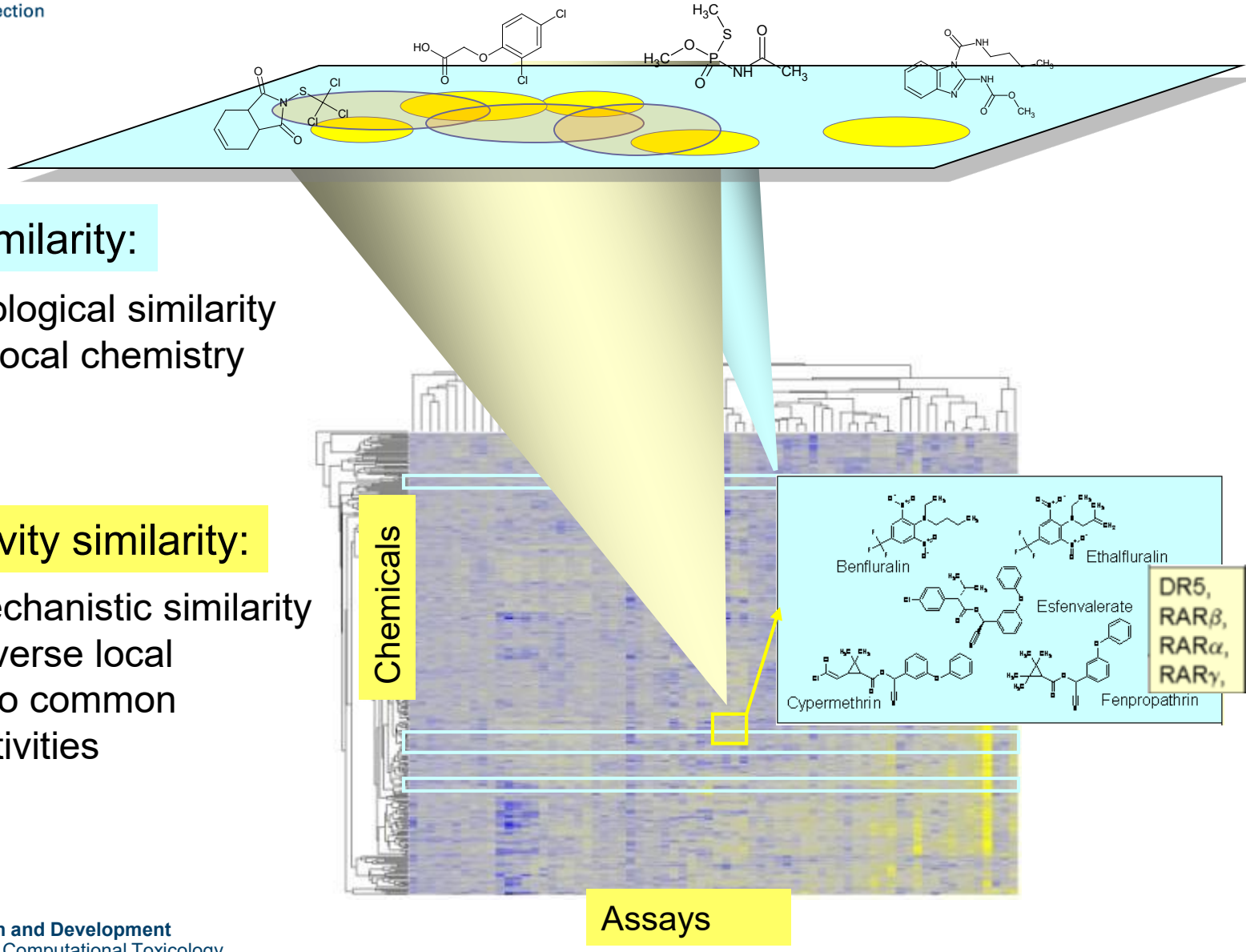


Predicting Toxicity: Toxicologist perspective



Structure vs. Bioactivity Similarity

United States Environmental Protection Agency



Structure similarity:

- implies biological similarity
- limited to local chemistry

HTS bioactivity similarity:

- implies mechanistic similarity
- can link diverse local chemistries to common biological activities

Toxicity Prediction Challenge

Biologically-based QSAR & Cheminformatics

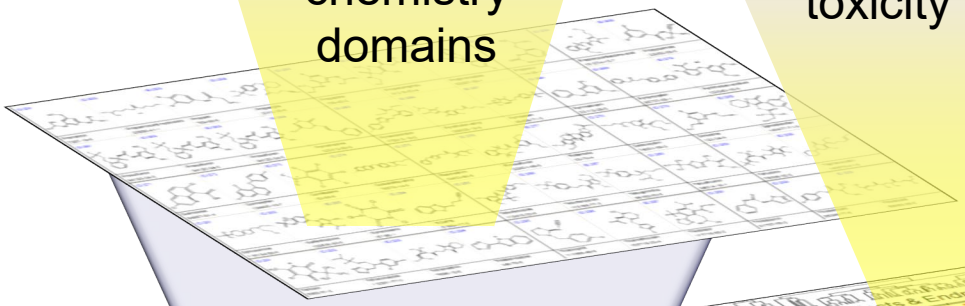
Reactivity & toxicity-informed features & local chemistry domains

Curation, aggregation

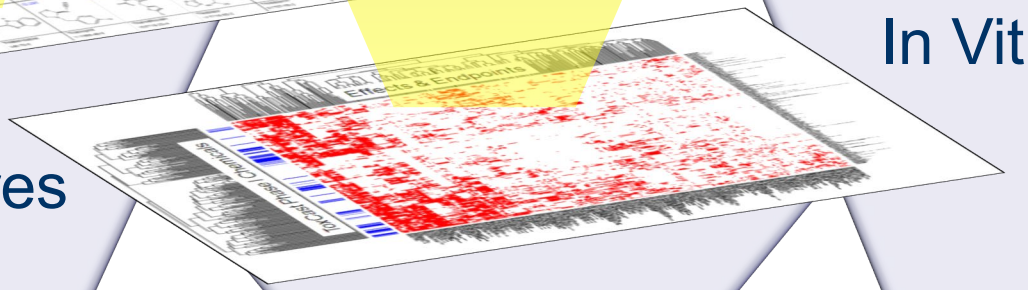
Model mechanistically well-defined toxicity endpoint

Data-mining

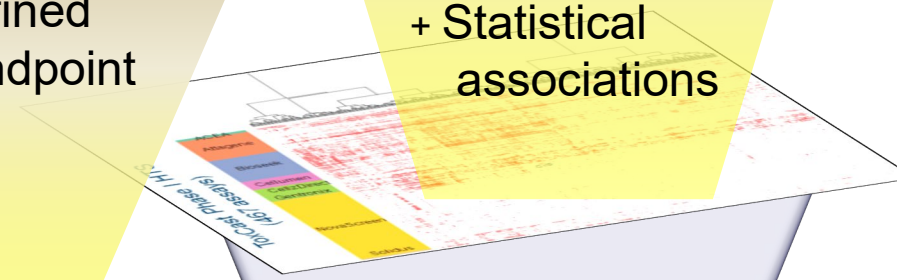
Adverse Outcomes:
> Pathways
> Genes
> Assays
+ Statistical associations



Structures



In Vivo

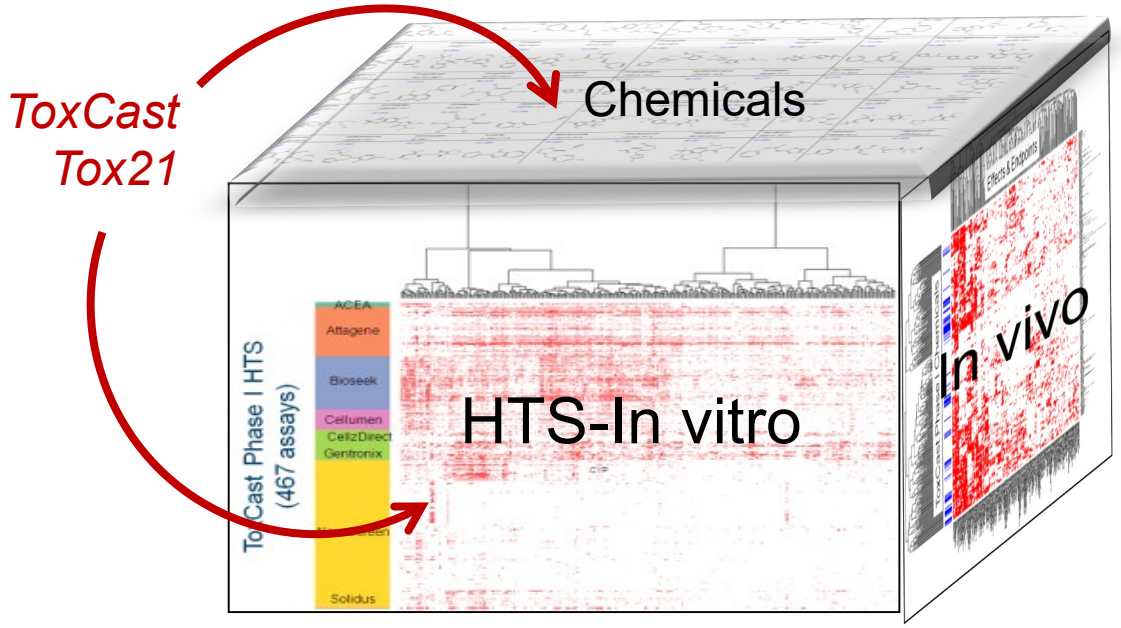


In Vitro/HTS

Existing knowledge

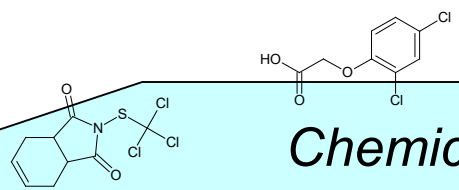
Chemical “probes” of biological activity

- Use existing knowledge & SAR to mine HTS data
- Use HTS data to inform & refine SAR models & approaches
- Use all of these data to improve ability to model toxicity

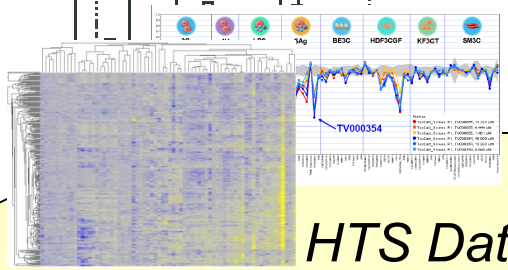


ToxCast: "Big Data" Informatics Challenges

United States Environmental Protection Agency



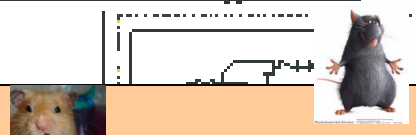
Chemical Structures



HTS Data

Biochemical, Cell-based, …

Standards
Ontologies
Databases
Informatics
User-interfaces
Public access

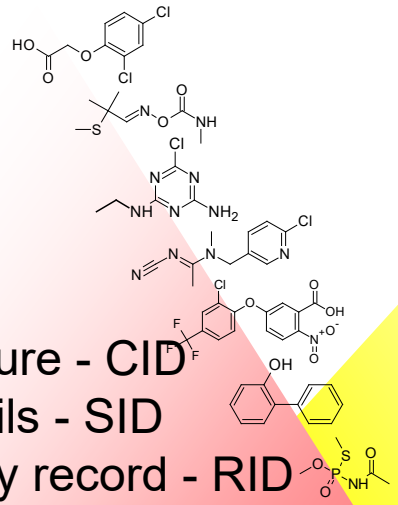


Bioassay Data

ToxRef DB, NTP, IRIS, etc

ToxCast/Tox21 Chemical Registry

United States Environmental Protection Agency



DSSTox

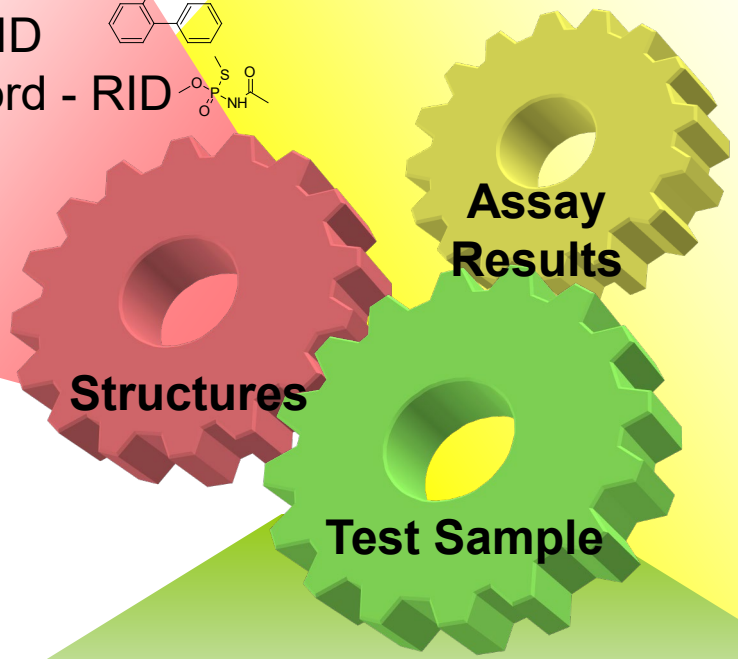
Chemical structure - CID
Substance details - SID
Project inventory record - RID

ACToR/InVitroDB

DSSTox RID

Assay name
Assay details
Assay outcome

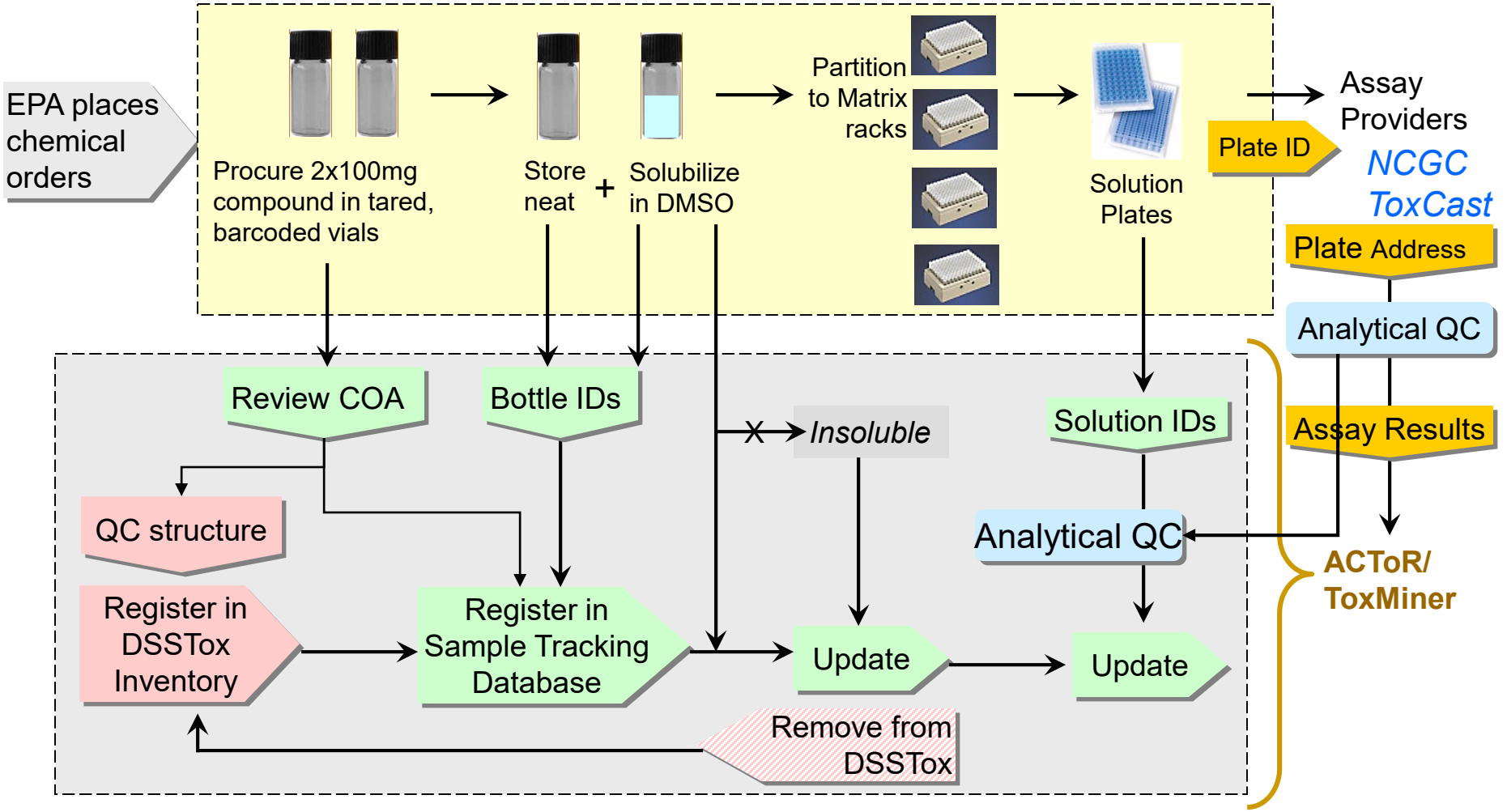
Solution ID
Plate ID
Plate Address ID



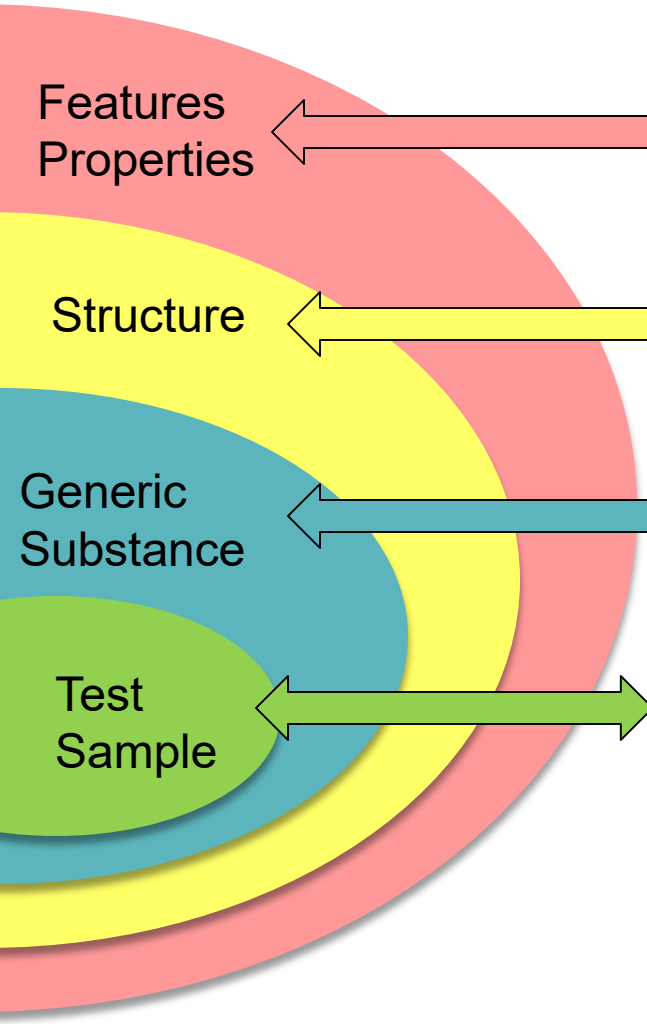
DSSTox RID
Bottle ID (→ COA ID)
Solution ID (→ QC ID)

ChemTrack Database

Chemical Sample Registration Workflow

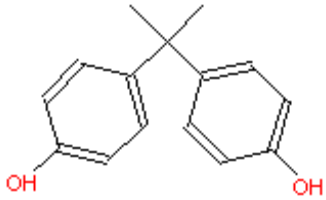


Chemical Representations supporting Data Integration



Chemotypes, fingerprints, phys-chem properties, ...

Chemical analogs, Read-across, SAR modeling



SMILES
InChI

Structure searching & modeling

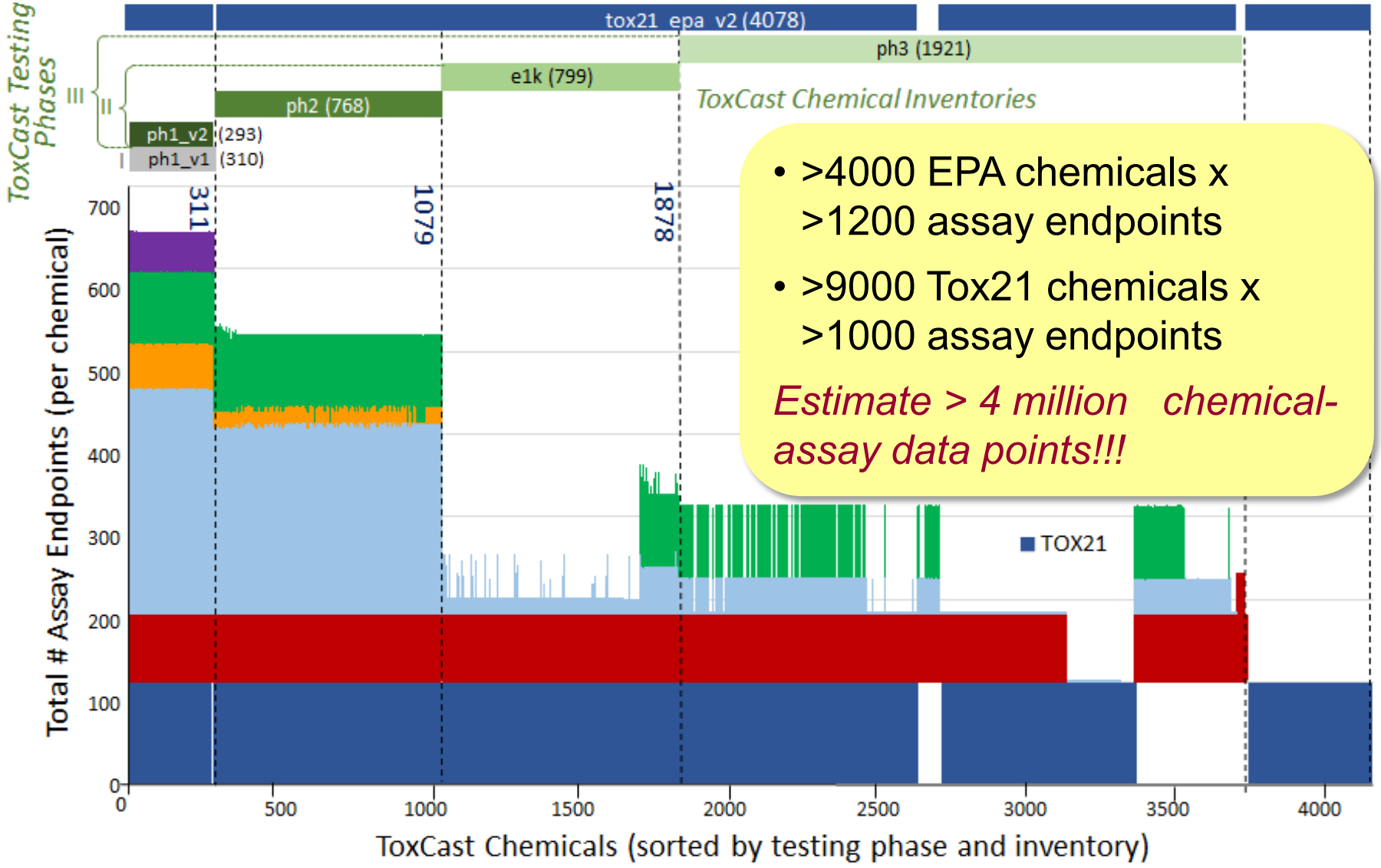
Chemical Name
CASRN

Public toxicity datasets

Supplier, Lot/Batch, physical description

Experimental Endpoint Data

ToxCast HTS data



The ToxCast library:

What's in it and is it fit for the purpose it was designed?

ToxCast Chemical Landscape

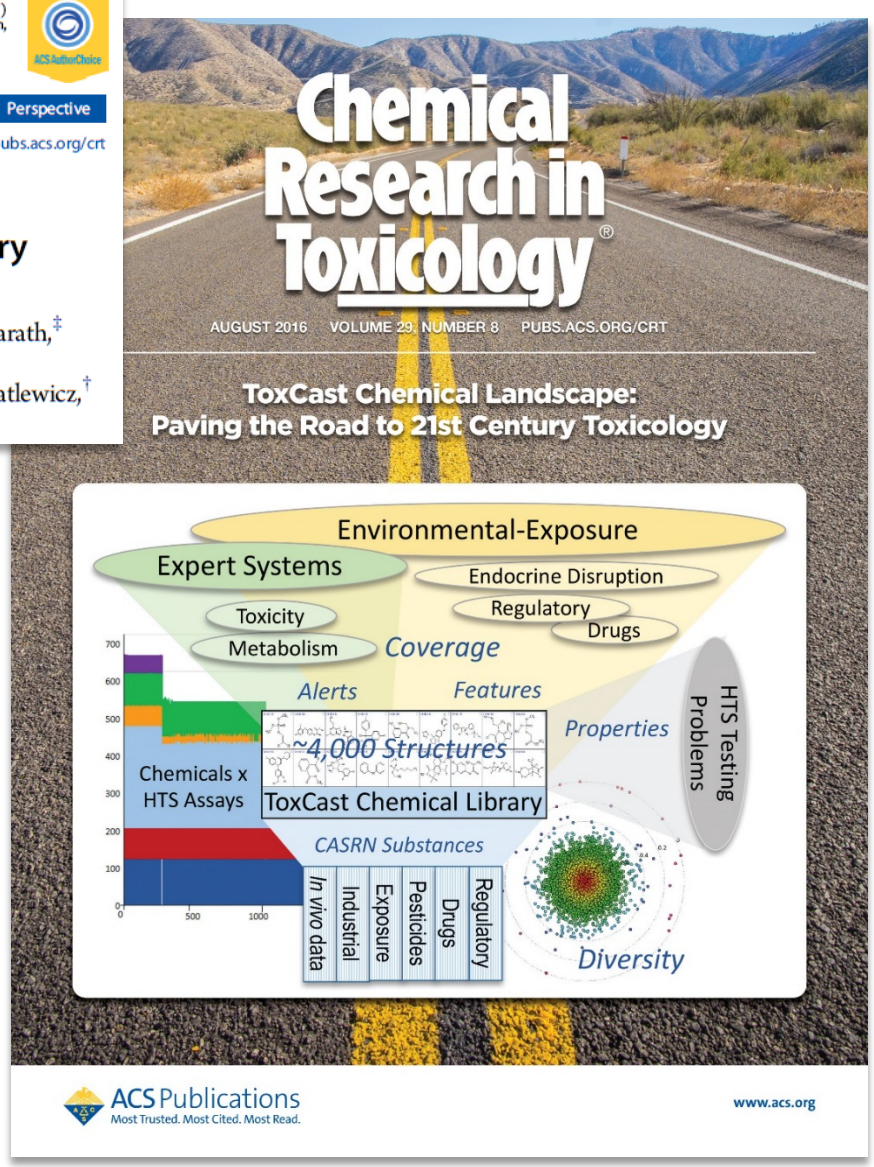
Chemical Research in Toxicology

Perspective

pubs.acs.org/crt

ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology

Ann M. Richard,*,† Richard S. Judson,† Keith A. Houck,† Christopher M. Grulke,† Patra Volarath,‡ Inthirany Thillainadarajah,§ Chihae Yang,∥,⊥ James Rathman,⊥,# Matthew T. Martin,† John F. Wambaugh,† Thomas B. Knudsen,† Jayaram Kancherla,▽ Kamel Mansouri,▽ Grace Patlewicz,† Antony J. Williams,† Stephen B. Little,† Kevin M. Crofton,† and Russell S. Thomas†



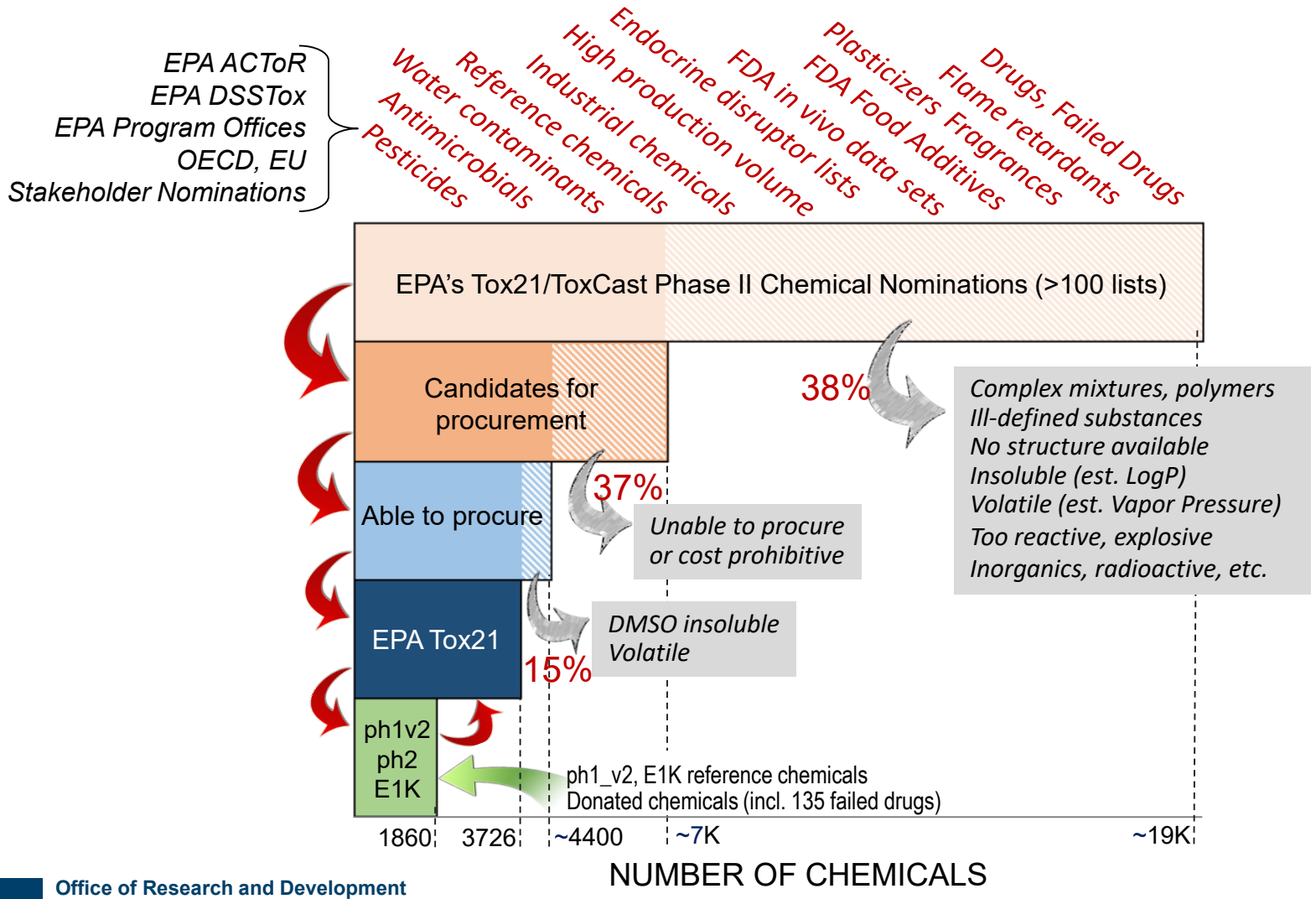
Open Access Perspectives article and Supporting Info files available for free download at:

http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00135

DOI: 10.1021/acs.chemrestox.6b00135

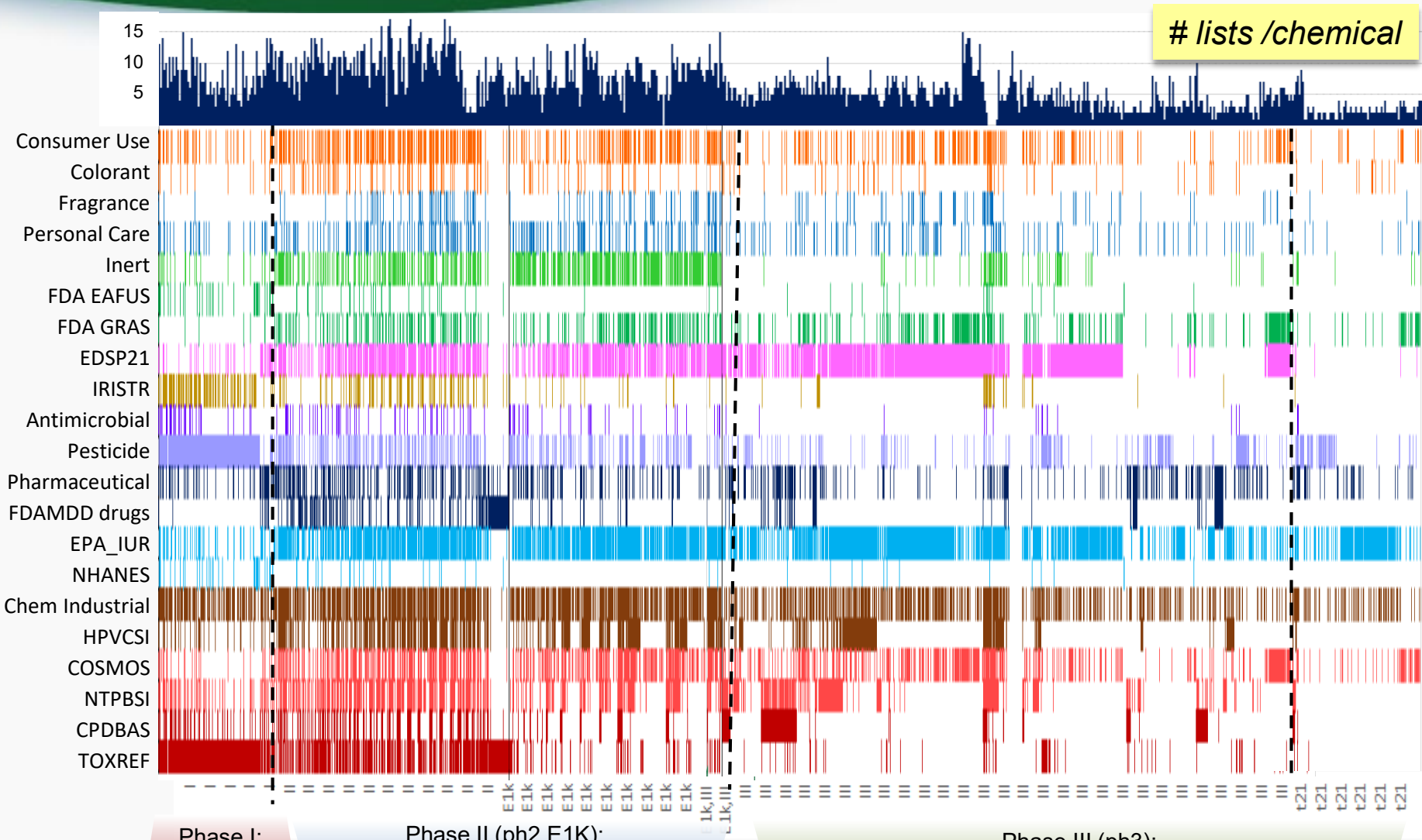
ChemResToxicol., 2016, 29, 1225−1251

Building EPA's ToxCast library



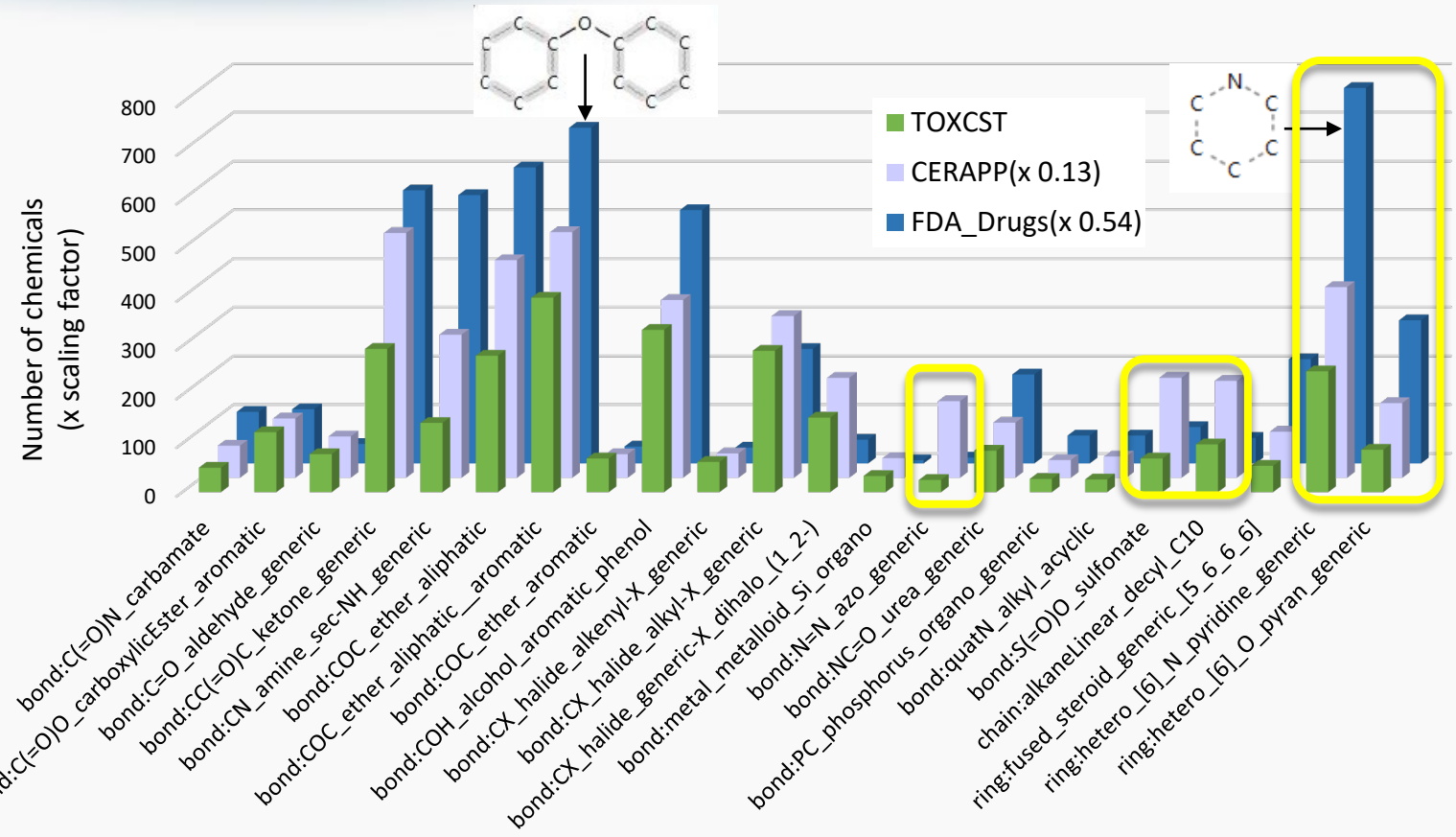
ToxCast Chemical Coverage: Use, Exposure, Toxicity

EPA United States Environmental Protection Agency



ToxPrint inventory profile comparisons (scaled)

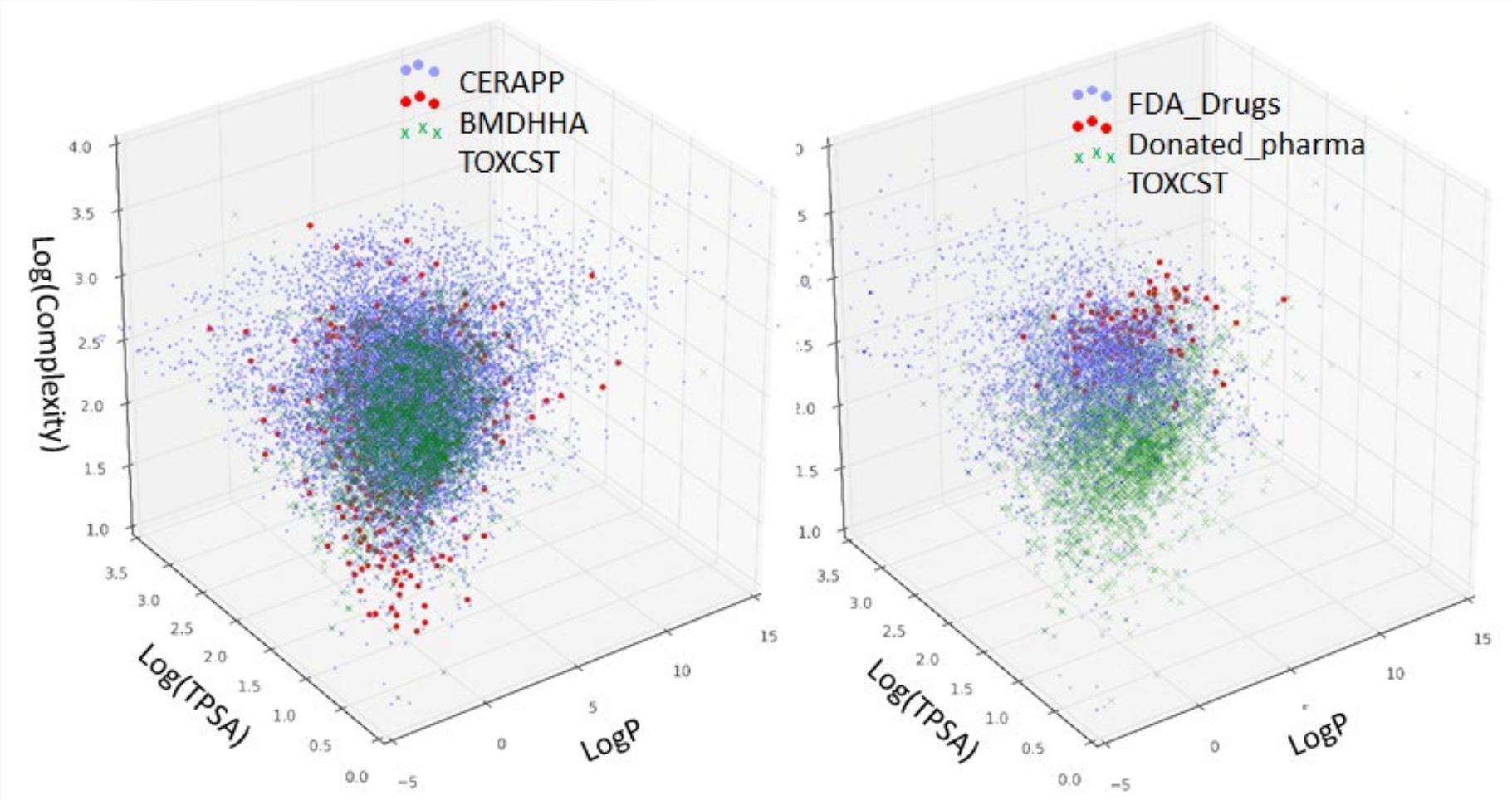
ChemResToxicol., 2016, 29, 1225−1251



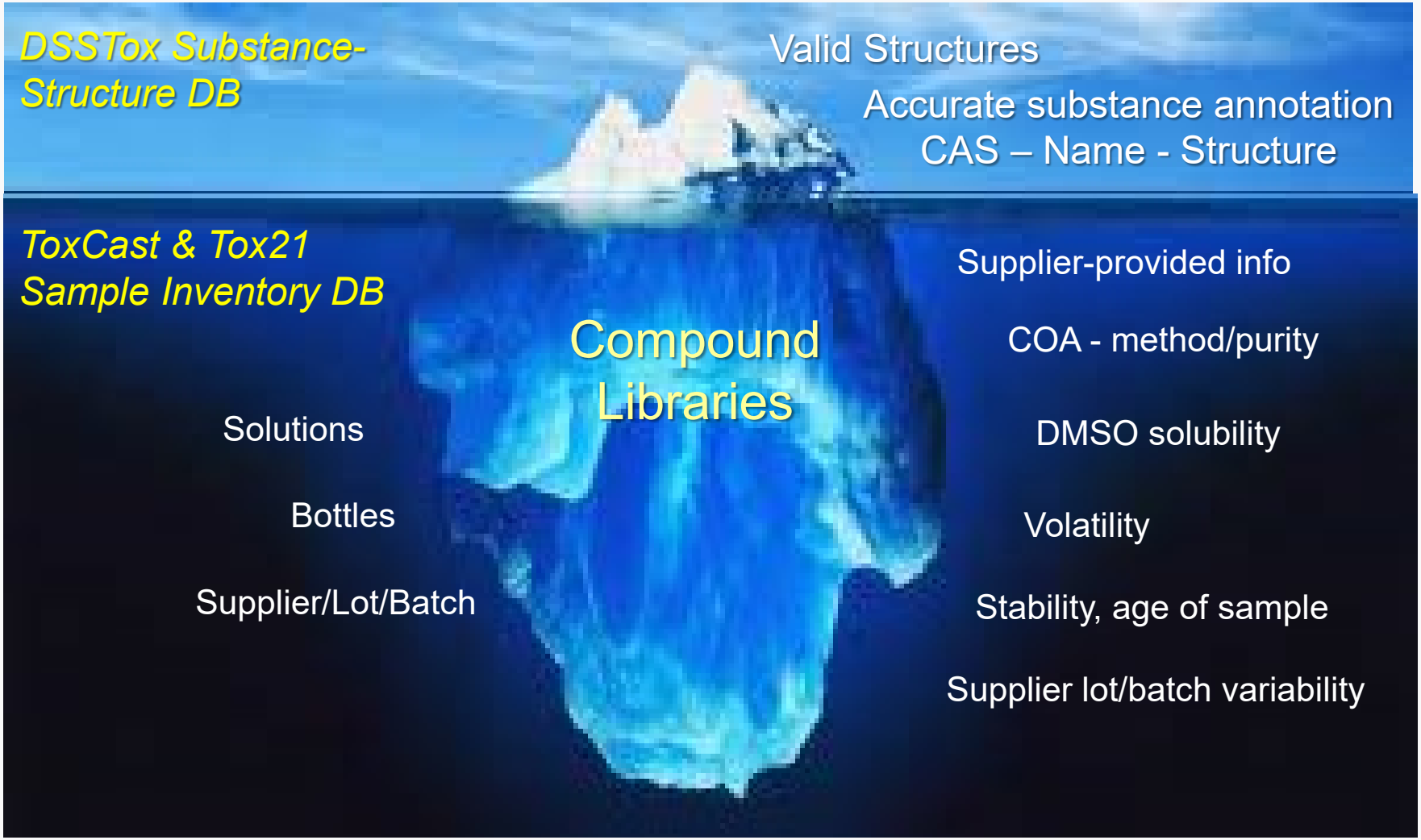
Comparison to potential target inventories based on computed properties

EPA United States Environmental Protection Agency

ChemResToxicol., 2016, 29, 1225−1251



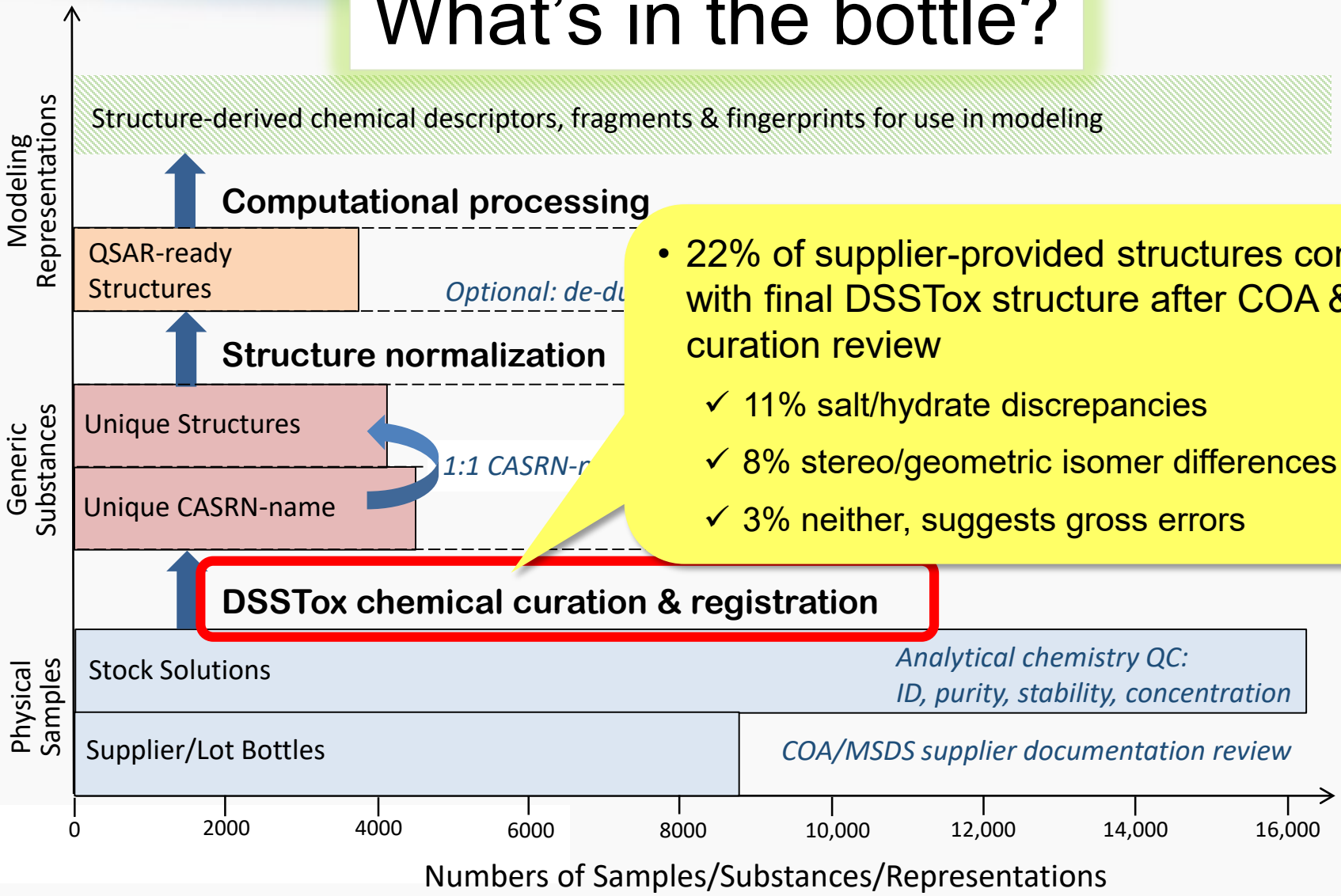
Chemical & Data Quality Issues



ToxCast Chemical Library: Quality Control Steps

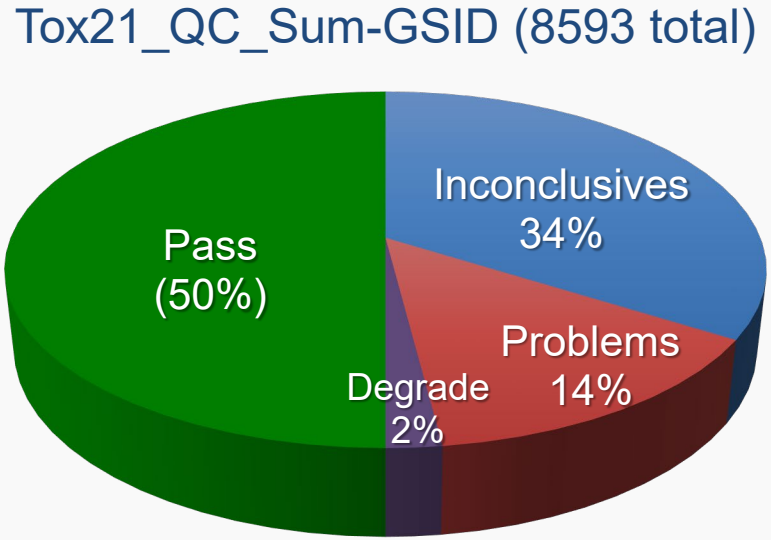
EPA United States Environmental Protection Agency

What's in the bottle?



- 22% of supplier-provided structures conflicted with final DSSTox structure after COA & curation review
  - ✓ 11% salt/hydrate discrepancies
  - ✓ 8% stereo/geometric isomer differences
  - ✓ 3% neither, suggests gross errors

Tox21 and ToxCast Chemical Library Analytical QC Results (8/2015)

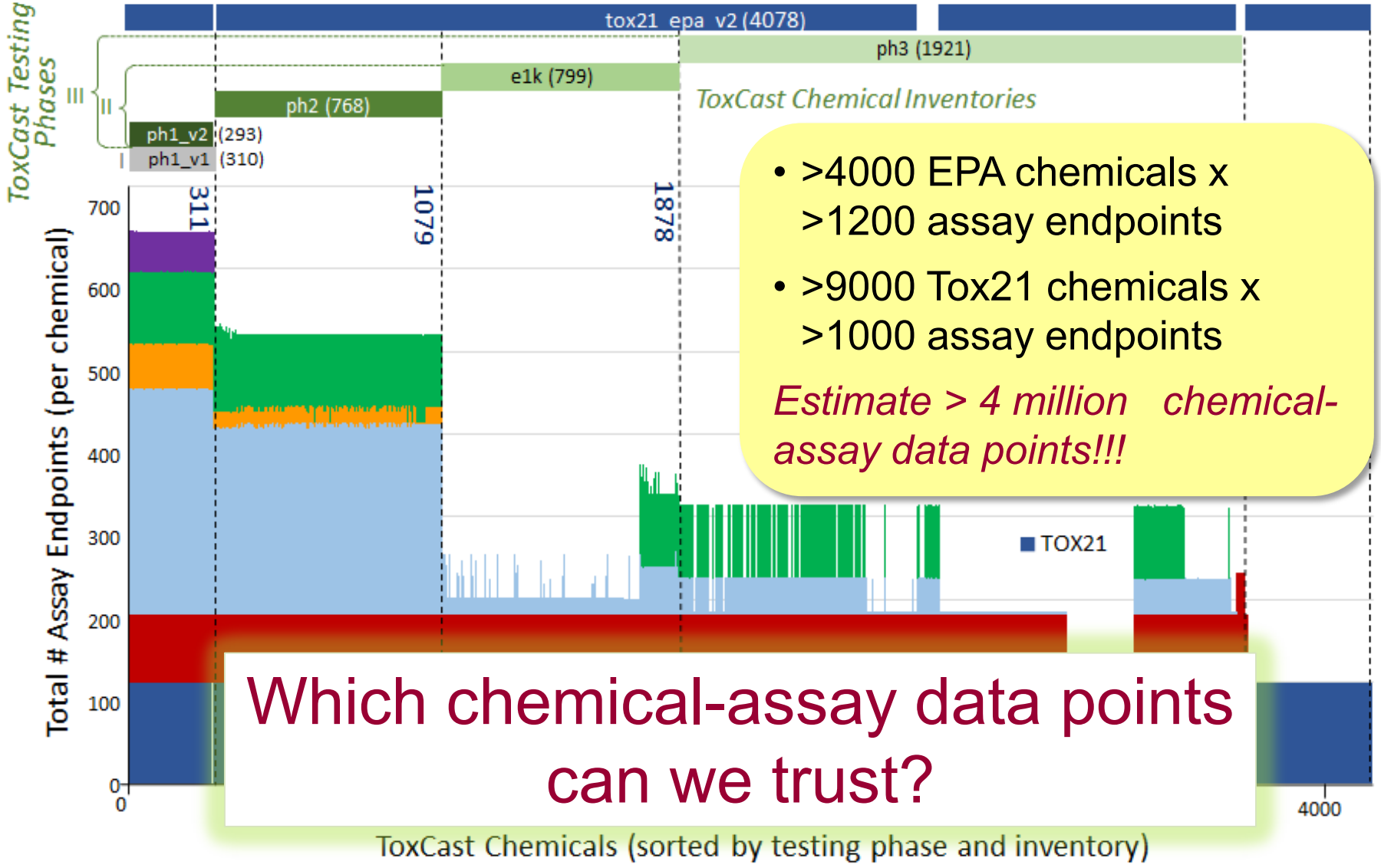


- 50% pass purity/ID/concentration checks
- A third(34%) of library pose analytical QC challenges (LCMS and GCMS only)
- 2% degrade after 4 months under testing conditions
- 14% problems - purity (<75%), ID and/or low concentration (<30% of expected [C])

- *Which chemicals have QC issues? (e.g., SVOCs?)*
- *Which chemicals were not analyzed? (e.g., mixtures, inorganics, etc.)*
- *How are HTS activity profiles linked to QC?*

Results as of 8/2015

ToxCast HTS data



- >4000 EPA chemicals x >1200 assay endpoints
- >9000 Tox21 chemicals x >1000 assay endpoints

Estimate > 4 million chemical-assay data points!!!

Which chemical-assay data points can we trust?

#4:
Importance of data quality

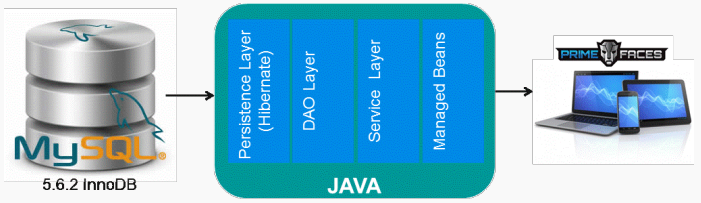
DSSTox Update

DSSTox_v1

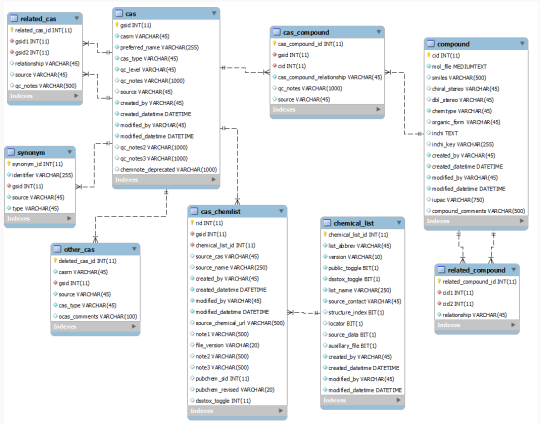
25,000 substances

DSSTox_v2

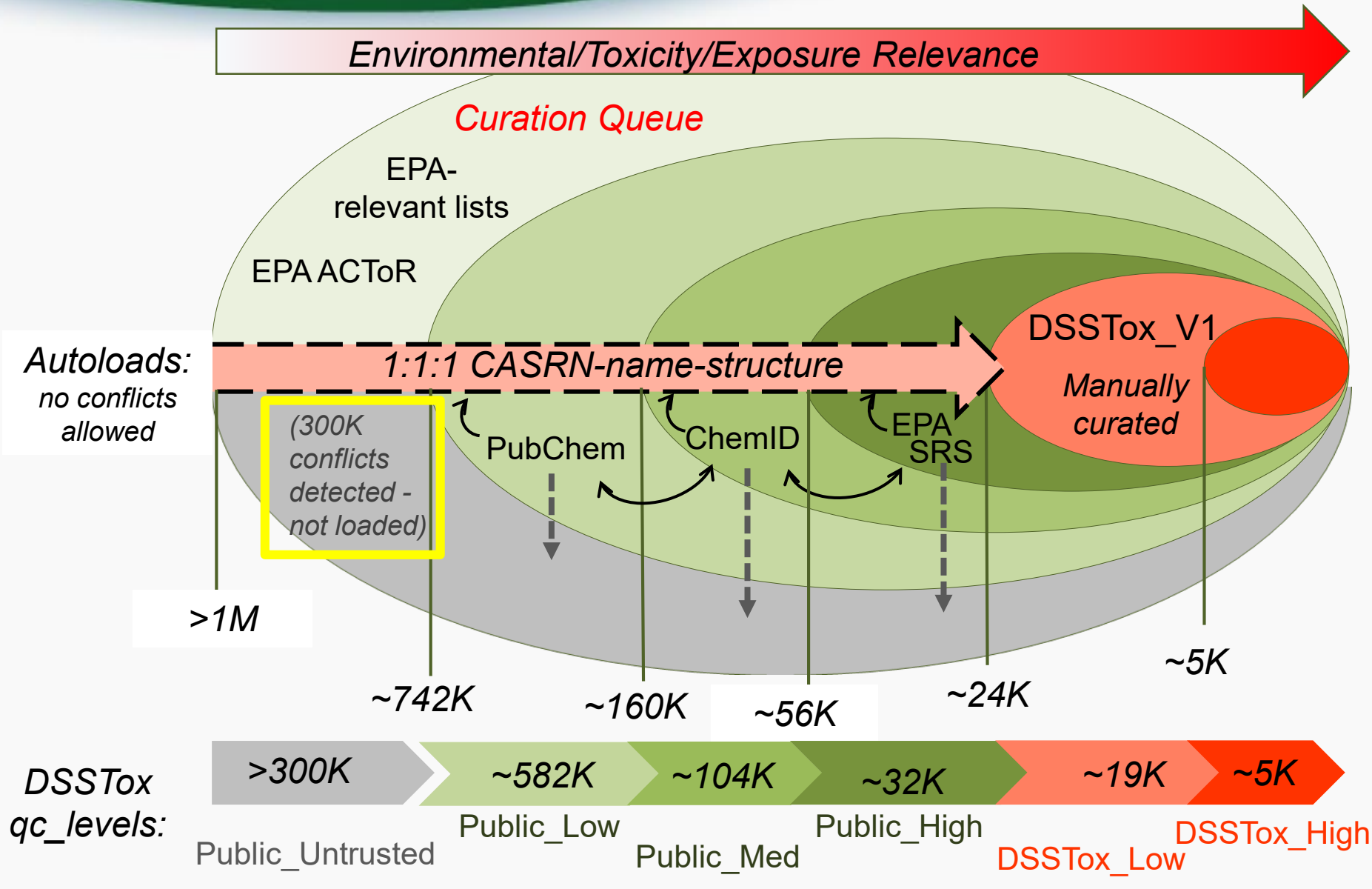
- Convert DSSTox tables to MySQL
- Develop curation interface & cheminformatics workflow
- Expand chemical content
- Web-services & Dashboard access



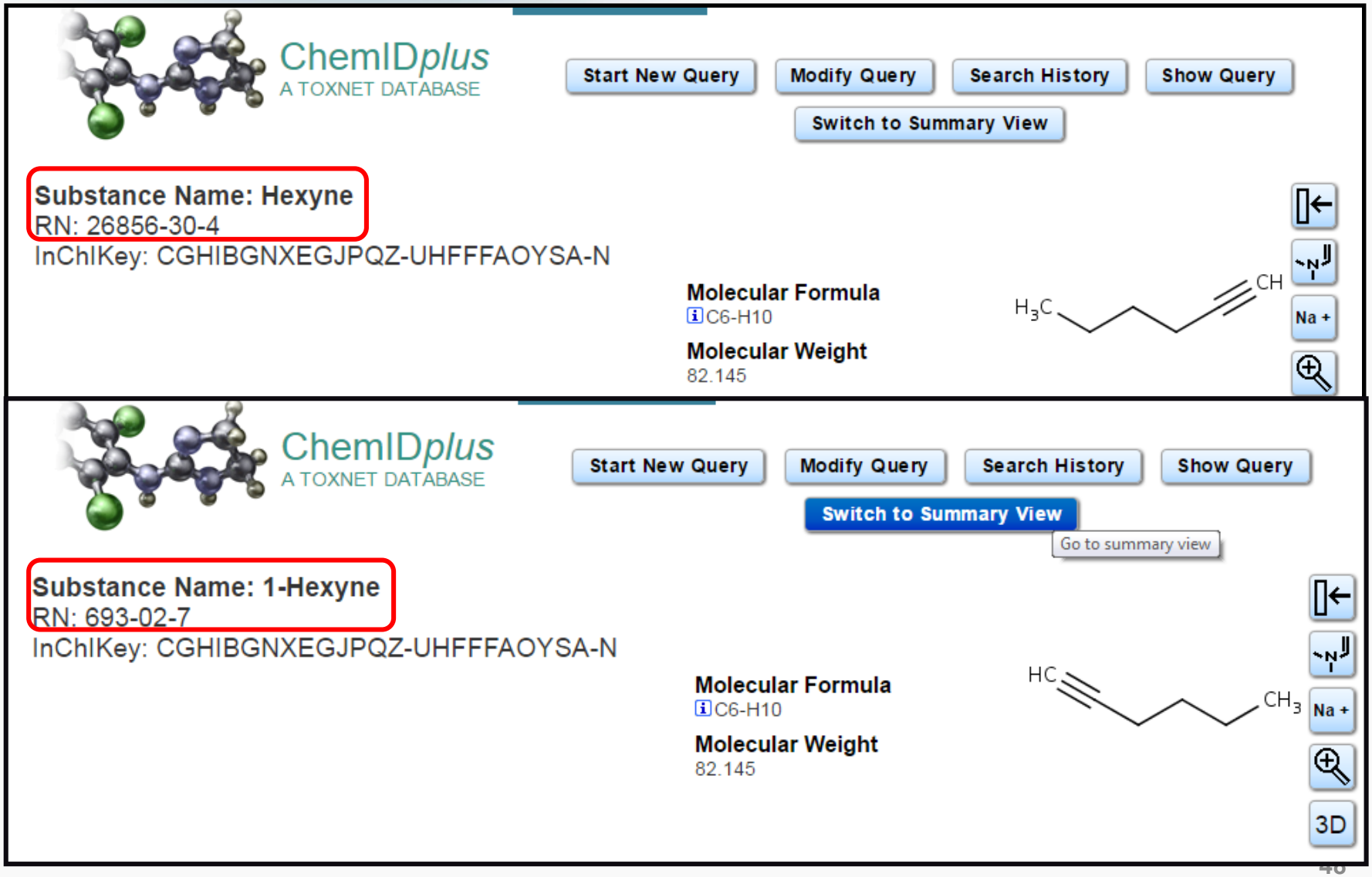
750,000 substances



Building DSSTox_v2

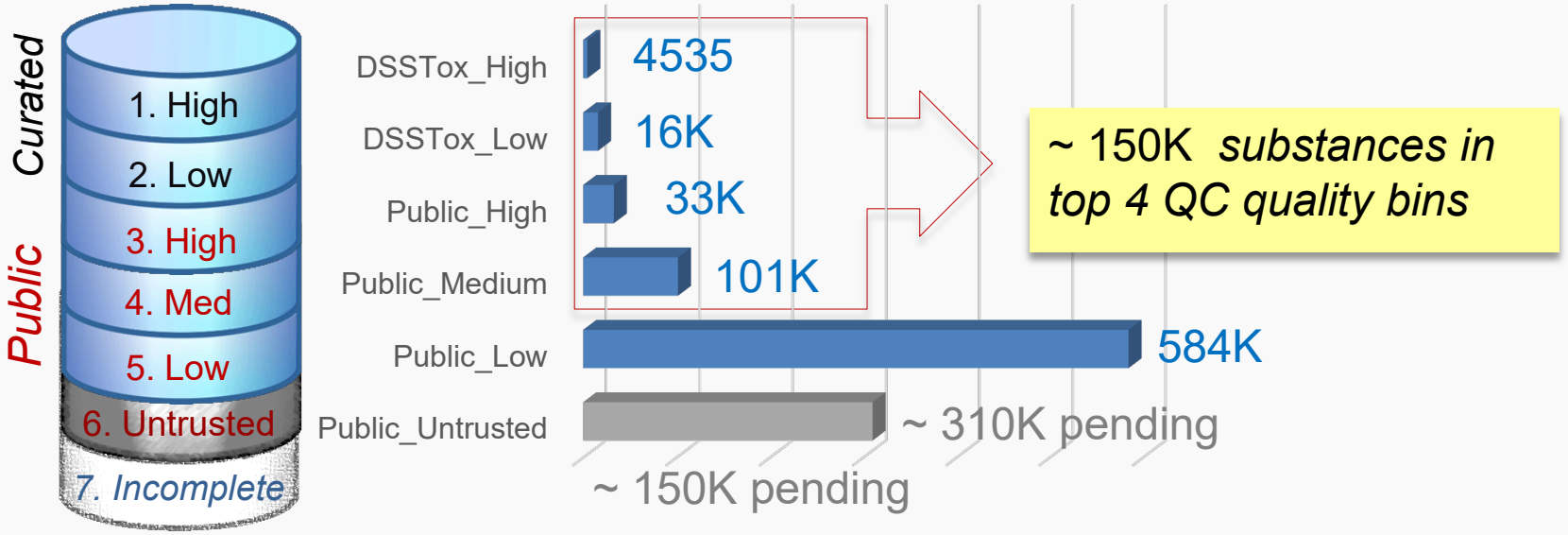


e.g., structure mapping collision



DSSTox_v2 Totals

QC Level Totals (12Jun2015)



QC Levels	
DSSTox_High:	Hand curated - highest confidence
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem (single source)
Public_Untrusted:	Postulated, but found to have conflicts in public sources

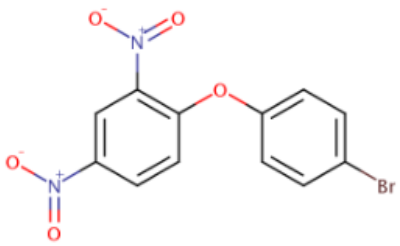
ChemReg Curation Interface

EPA United States Environmental Protection Agency

View/Edit a Single Record | Structure Search | Browse/Curate Records | Export DSSTox | Chemotypes | Login

CAS-RN matched
null
You are viewing the record associated with
DTXSID0022270
CASRN: 17589-66-1

17589-66-1



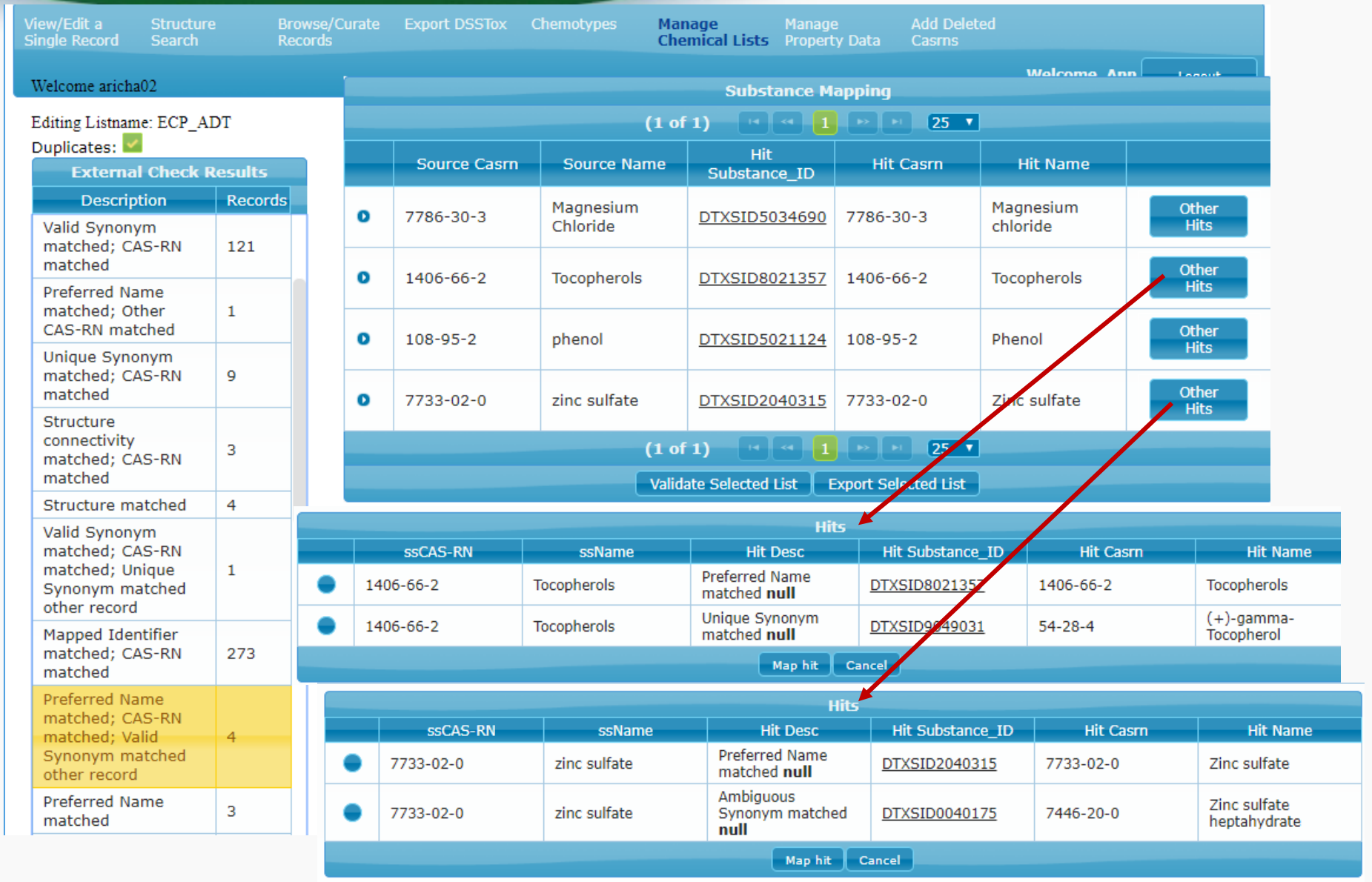
Systematic Name: 1-(4-Bromophenoxy)-2,4-dinitrobenzene
MolFormula: C12H7BrN2O5
InChI Key: BJGBENCKKXEWGB-UHFFFAOYSA-N
Smiles: [O-][N+](=O)C1=CC(=C(OC2=CC=C(Br)C=C2)C=C1)[N+]([O-])=O
PubChem ID: 221811
Chemspider ID: 192492

Substance ID: DTXSID0022270
CAS: 17589-66-1
Name: 4-Bromo-2',4'-dinitrodiphenyl ether
Substance Type: Single Compound
QC Level:
Data Source: Public
QC Notes:

Compound ID: DTXCID702270
Chemical Shown: Tested Chemical
Internal QC Notes:
Source of CAS-Compound: Public
Double Stereo:
Chiral Stereo:
Chemical Form: Organic
Organic Form:Parent

- Associated Lists (10)
- Synonyms (0)
- Other Cas (0)
- Successor Substances (0)
- Predessessor Substances (0)

List Curation Interface

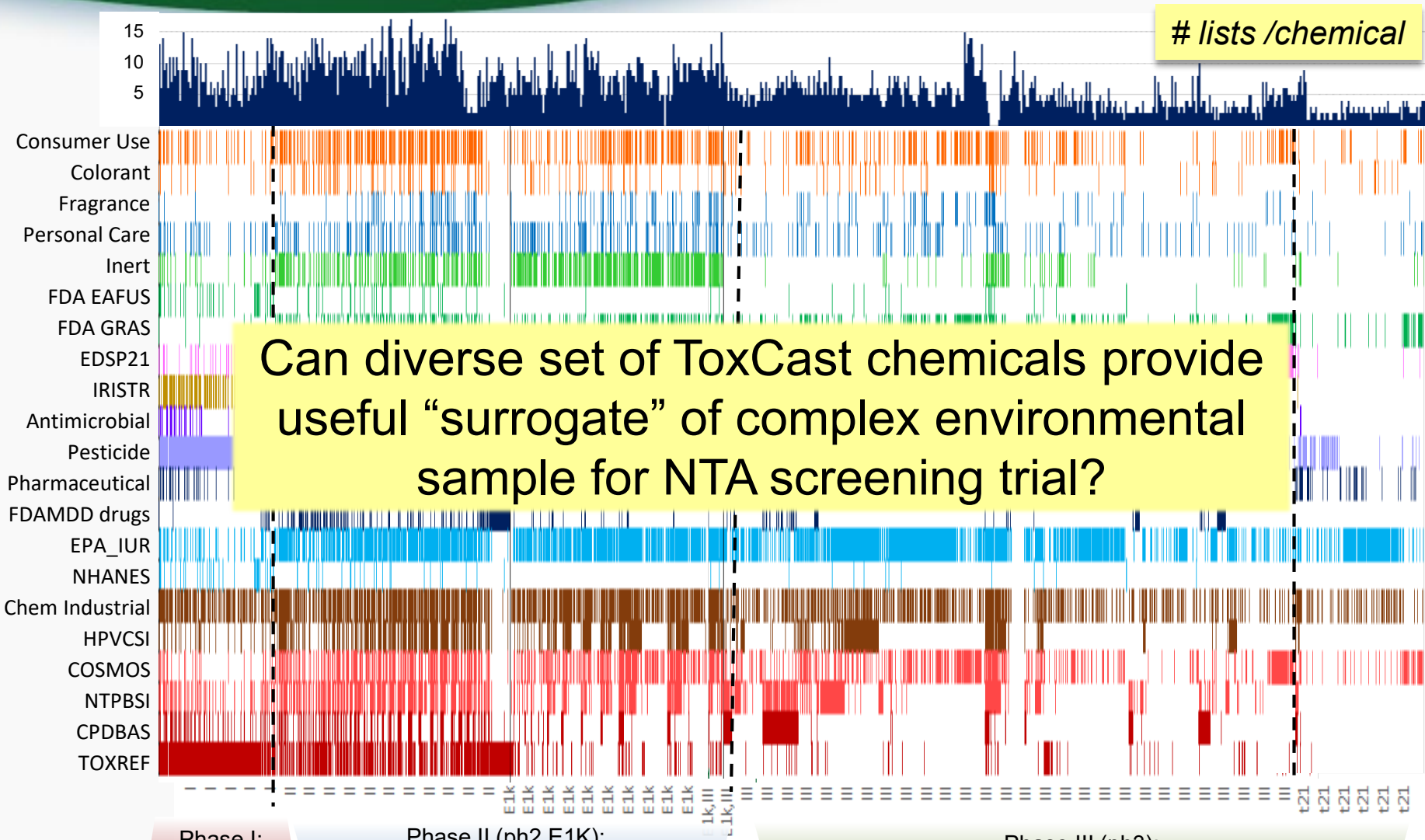


#5:
Importance of data quality

Now that we've spent 10 years building the ToxCast sample library, how else can we use it?

ToxCast Chemical Coverage: Use, Exposure, Toxicity

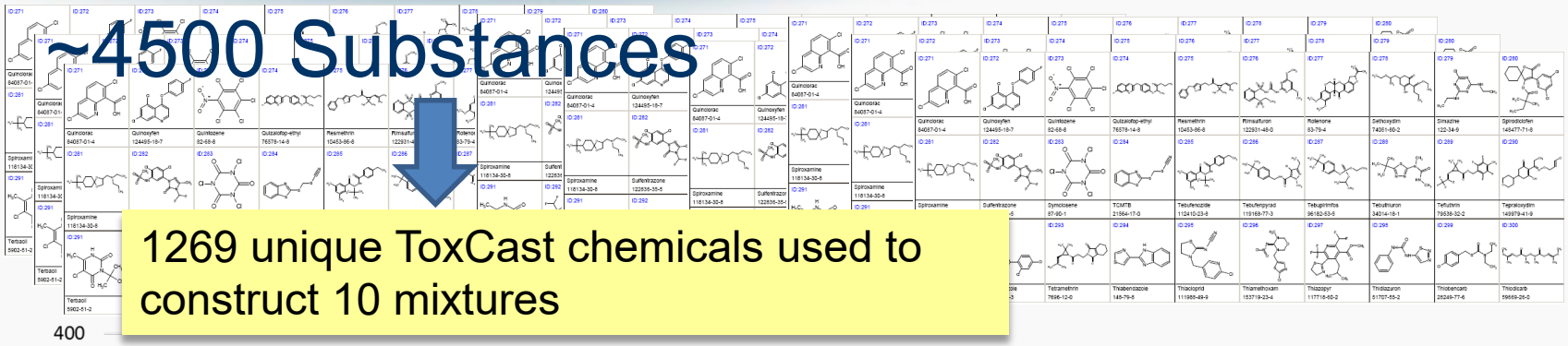
EPA United States Environmental Protection Agency



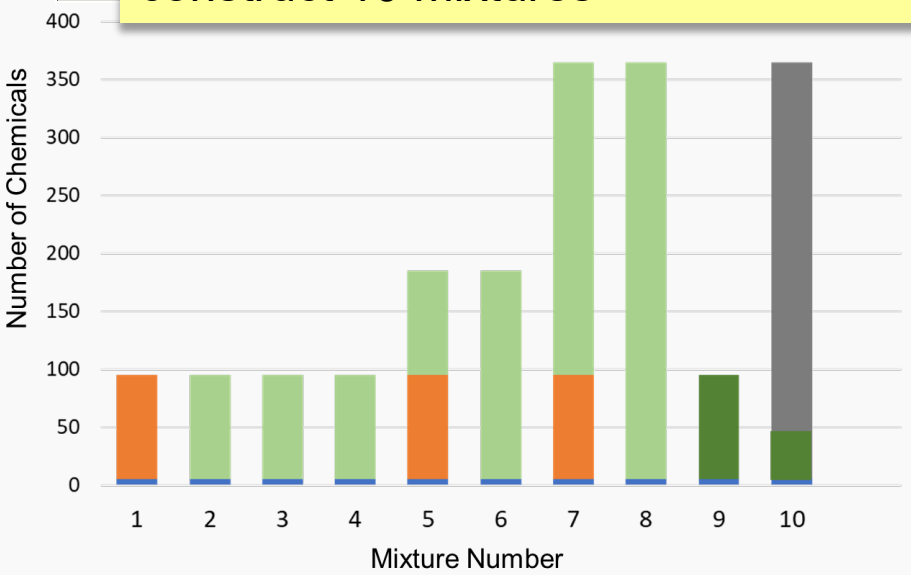
ENTACT Mixture Trial

EPA United States Environmental Protection Agency

~4500 Substances



1269 unique ToxCast chemicals used to construct 10 mixtures

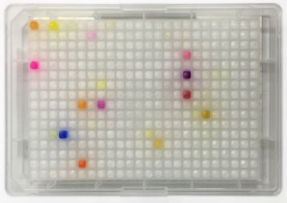


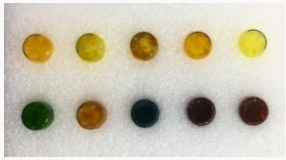
Chemicals in Mix 1-8 (amenable):

- *"Grade A" analytical QC results (LC/GC only)*
- *single DSSTox structure*

Chemicals in Mix 9,10 (challenging):

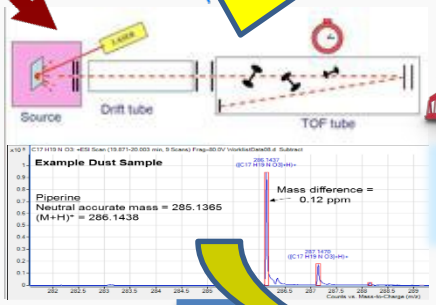
- *contain isomeric & isobaric cmpds*
- *contain cmpds graded as <80% purity*





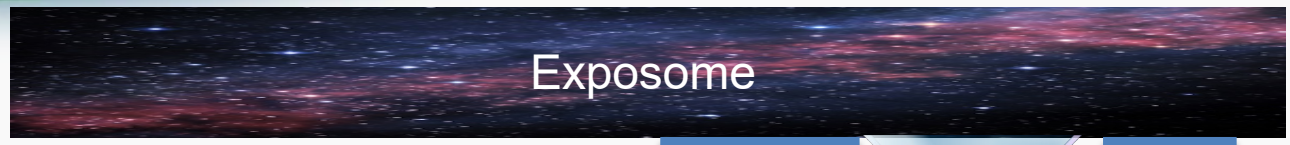
Cheminformatics view of non-targeted testing problem

EPA United States Environmental Protection Agency



Thousands of mass ion peaks & abundances

What's in the sample?
&
Should I be worried?



Exposure?

Toxic?

Suspect testing,

ENTACT

{mono-isotopic mass}

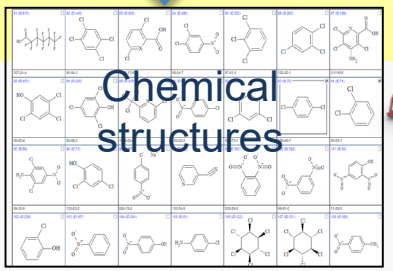
MI mass

1:many

{formulas}

1:many

{MS-ready structures}



CID structure

DSSTox chemical substance database

Inventory/Data

CURATION

{CAS-name}

CURATION

1:1

{structure}

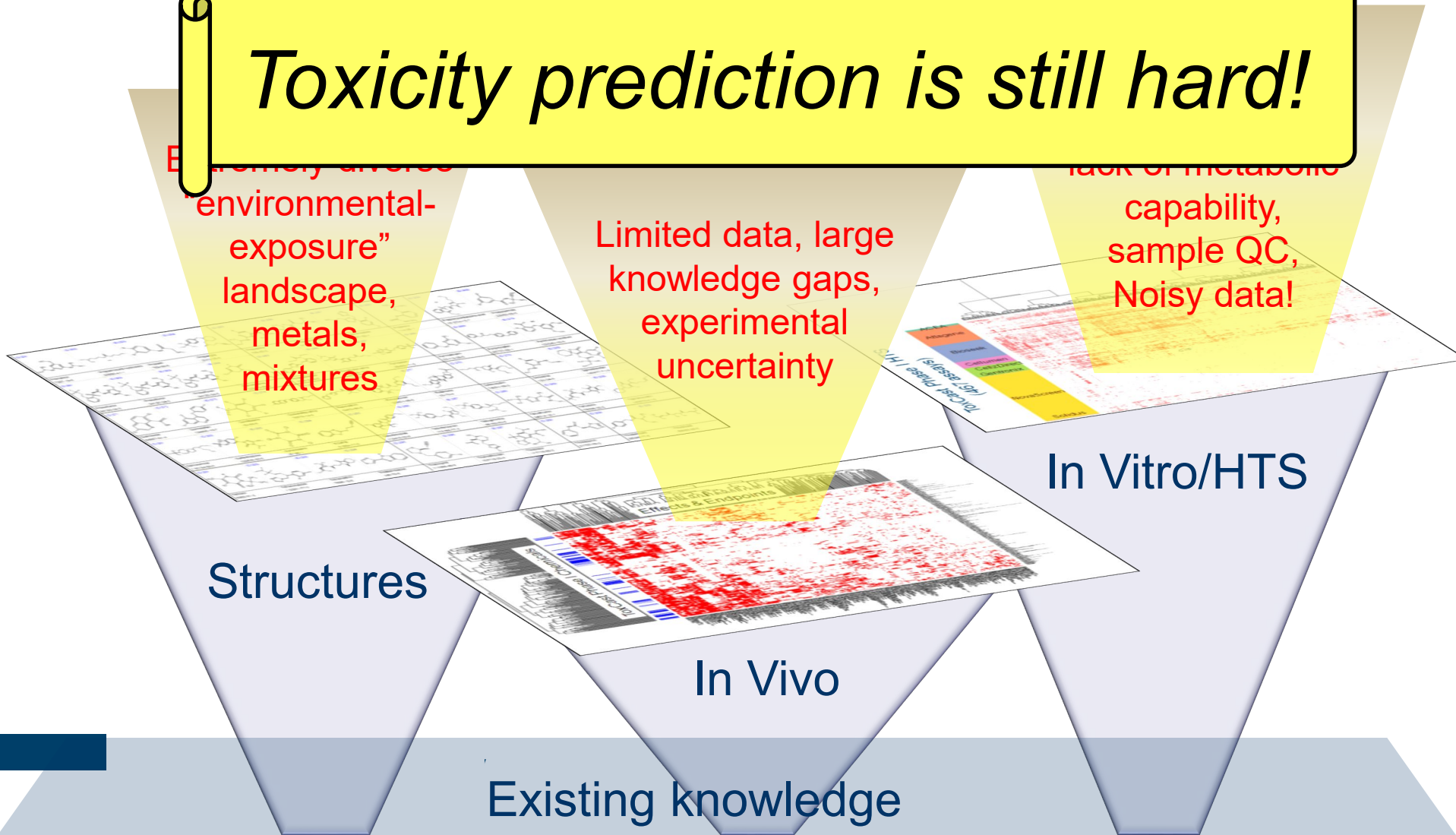
CURATION

#6:
Controlling a sample library and data foundation grants

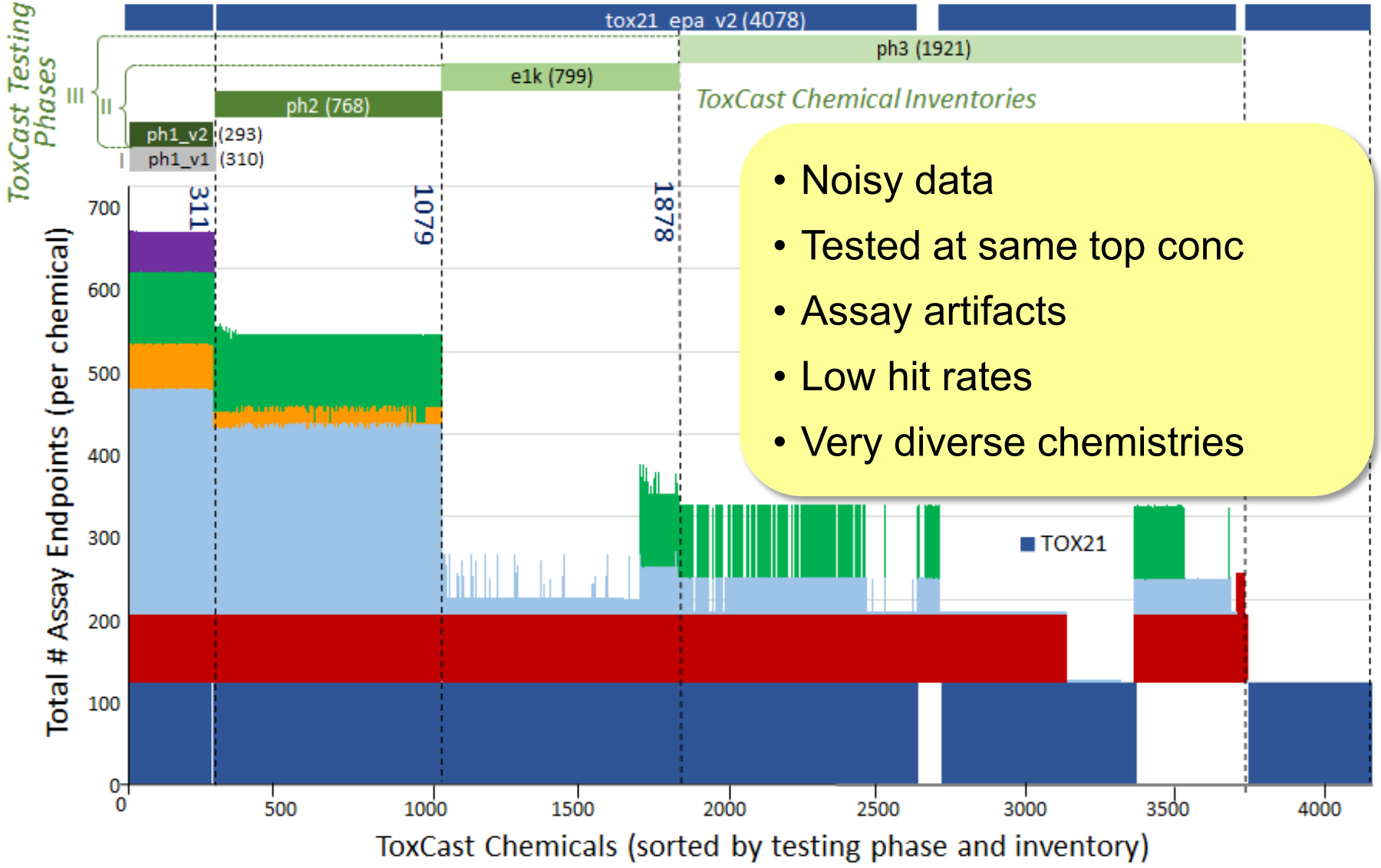
Returning to the challenge of mining our data and predicting toxicity…

Toxicity Prediction Challenge

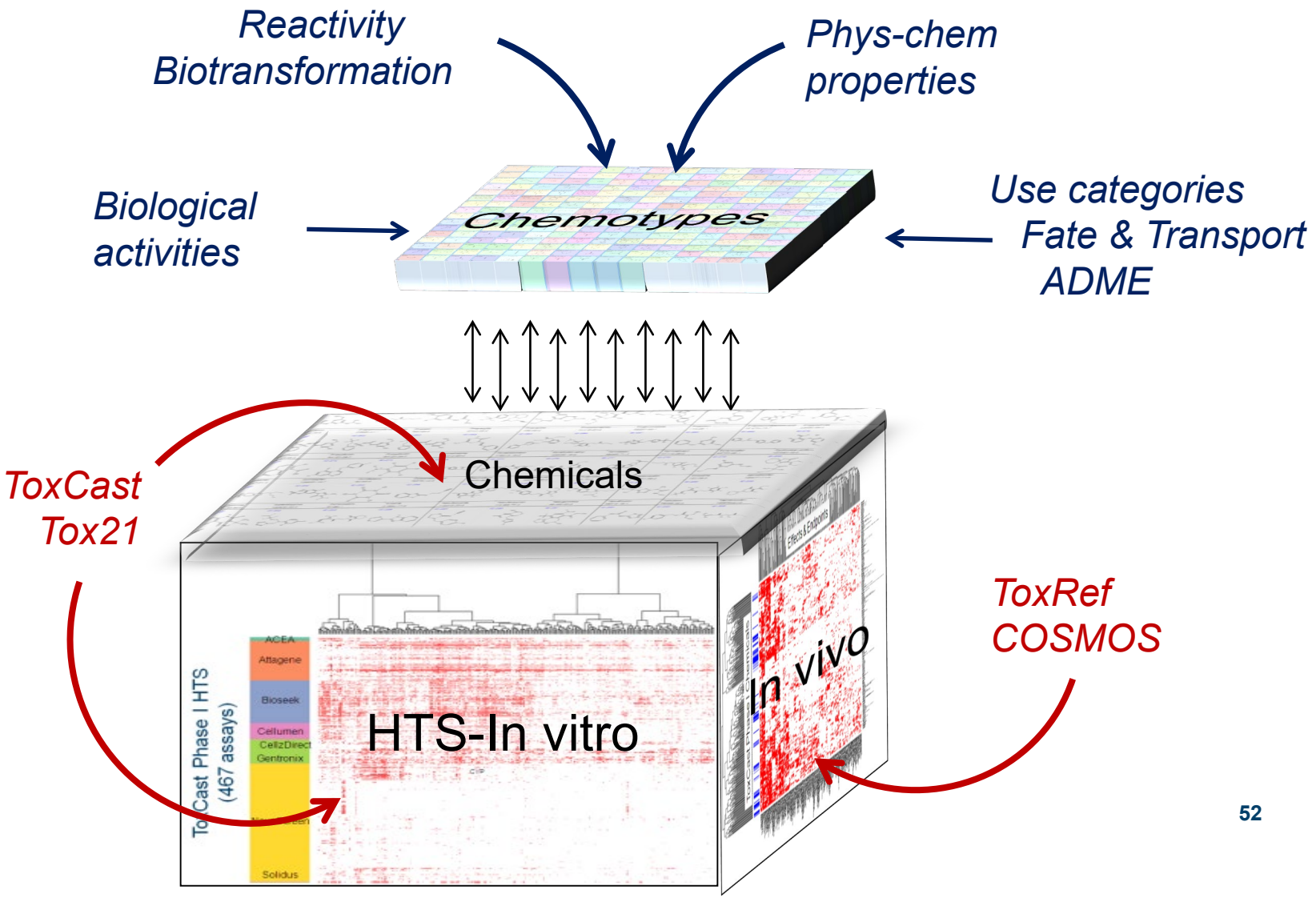
Toxicity prediction is still hard!



ToxCast HTS data



Building a public chemotype "knowledge- base"



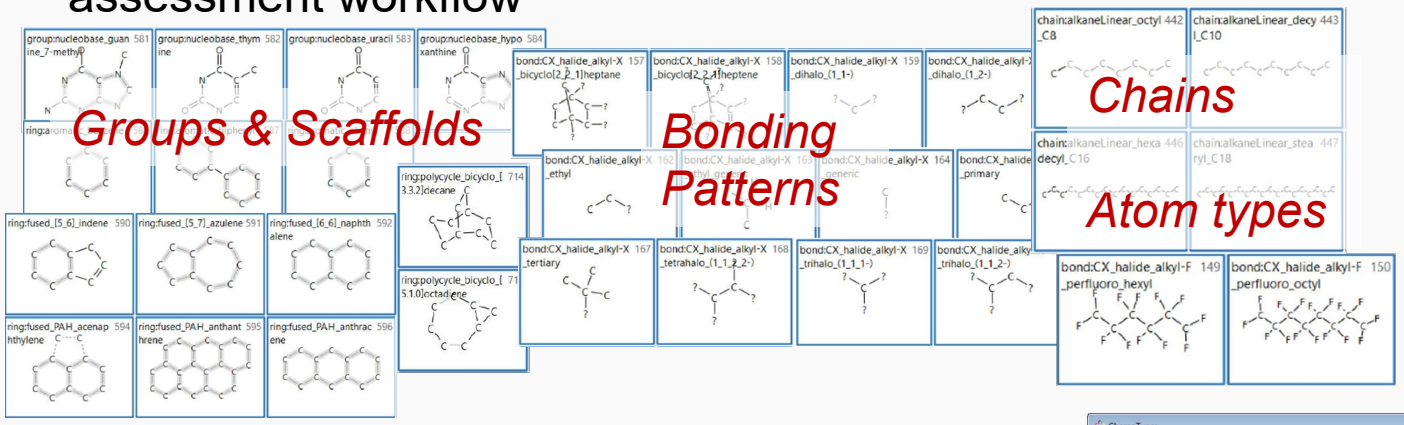
ToxPrints: A Public Set of Chemotypes

EPA United States Environmental Protection Agency

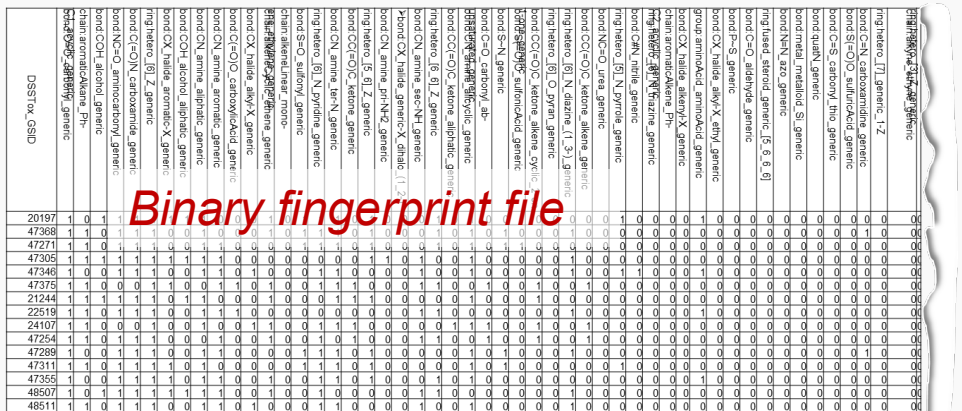
→ *Clear, reproducible means for defining regions of local chemistry.*

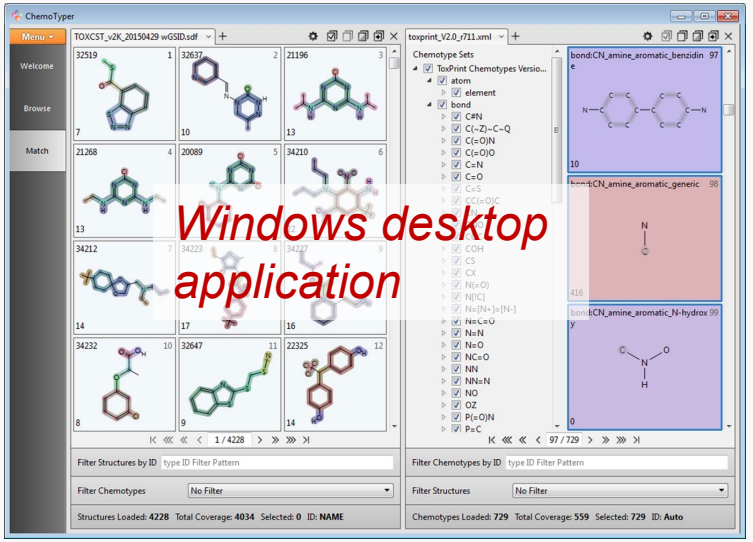
ToxPrints: http://www.toxprint.org

729 features important to EPA & FDA's "chemical exposure" landscape and safety assessment workflow

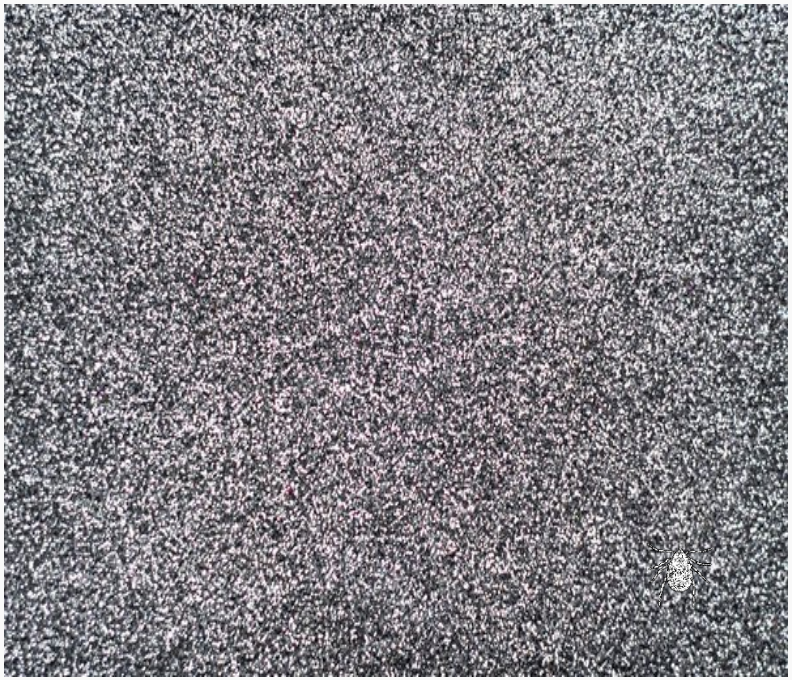


Chemotyper: http://www.chemotyper.org

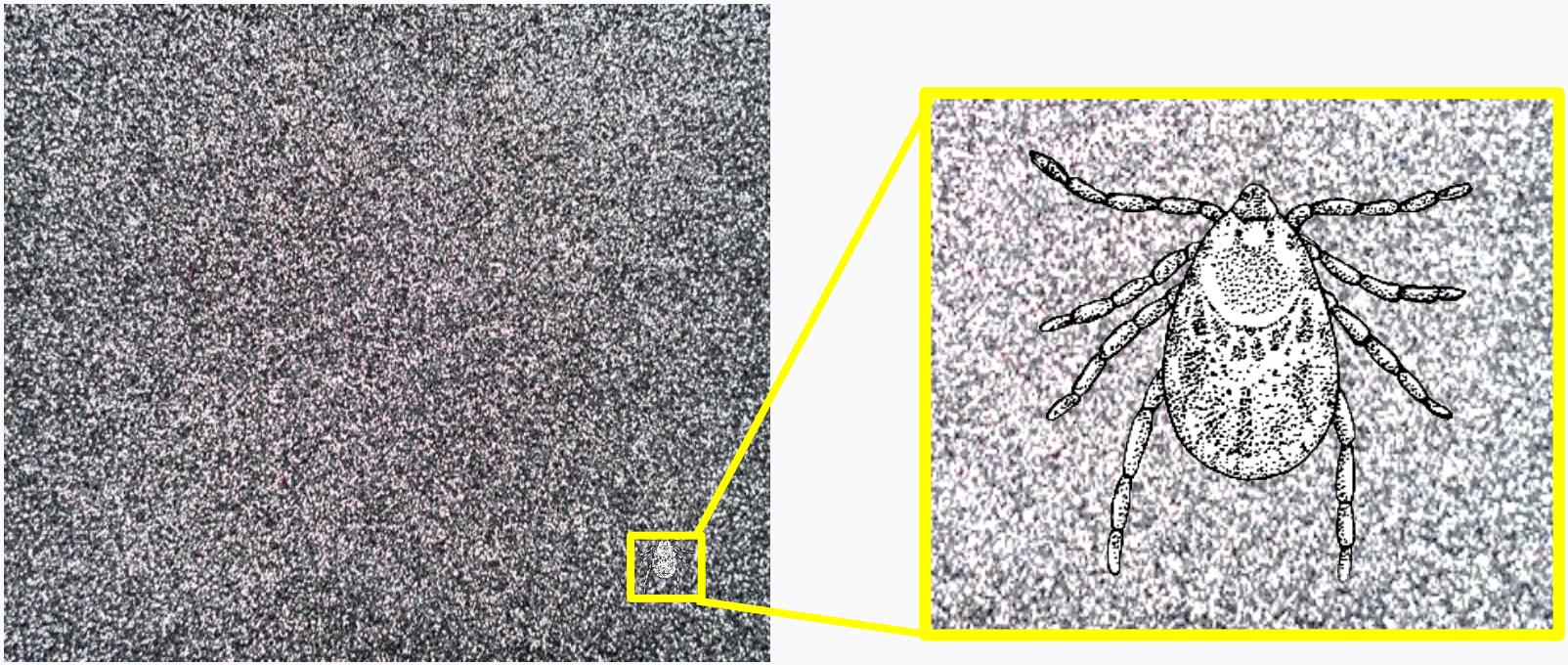




Concept of Enrichment:
Focus & Amplify to See



Concept of Enrichment:
Focus & Amplify to See



Chemotype Enrichment,
e.g. *Flame Retardant (FR) Use Category*

- 42 FRs amenable to HTS included in ToxCast Ph3 & Tox21

$\{CT\}_{FR}$

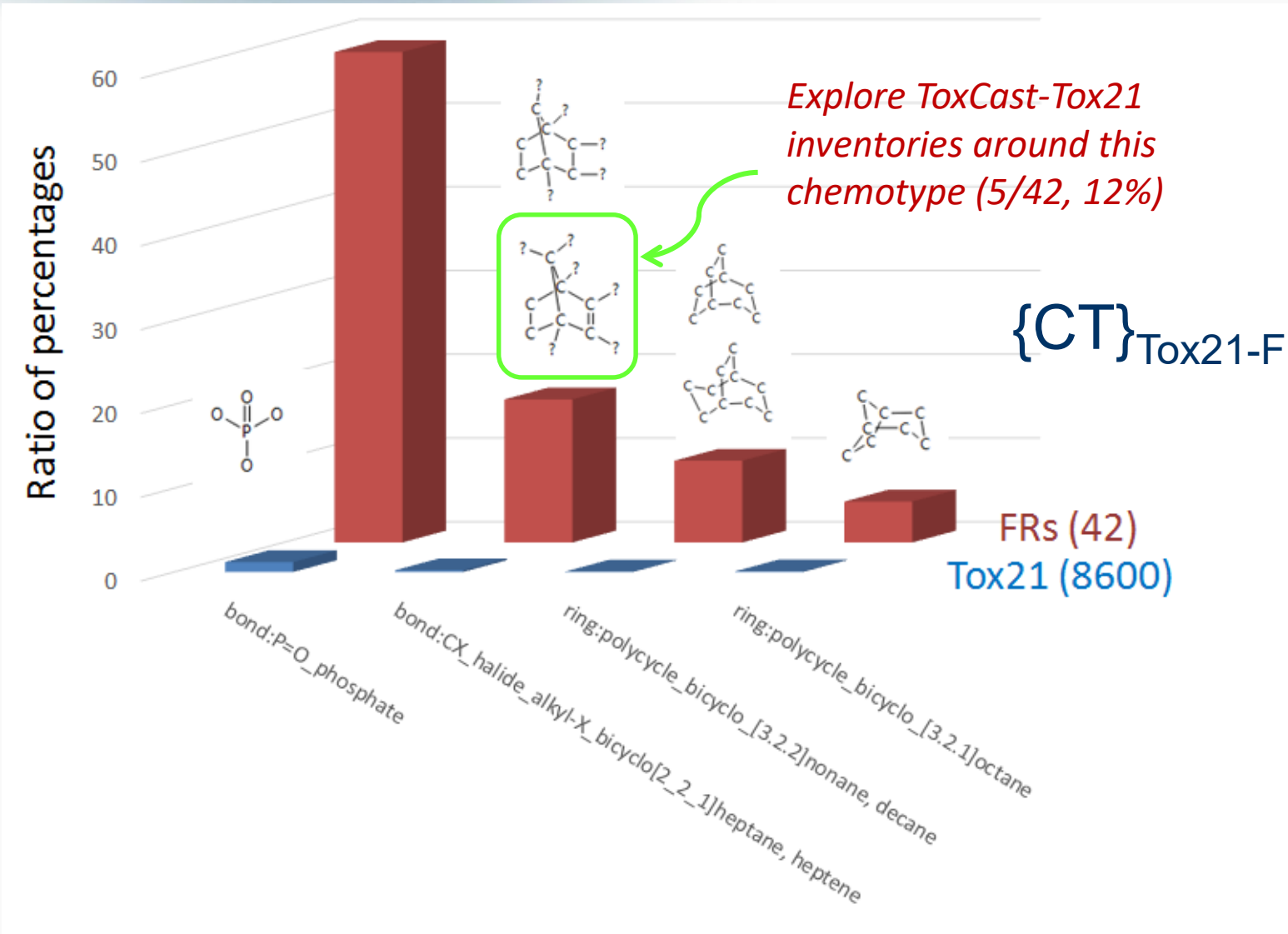
3296-90-0	2,2-Bis(bromomethyl)-1,3-propanediol
115-28-6	Chlorendic acid
2921-88-2	Chlorpyrifos
2385-85-5	Mirex
115-96-8	Tris(2-chloroethyl) phosphate
78-42-2	Tris(2-ethylhexyl) phosphate
115-86-6	Triphenyl phosphate
126-73-8	Tributyl phospha
79-94-7	3,3',5,5'-Tetrabro
13674-87-8	Tris(1,3-dichloro
1163-19-5	Decabromodiph
19660-16-3	2,3-Dibromoprop
563-04-2	Tri-m-cresyl pho
20120-33-6	Phosphonic acid dimethyl ester
868-85-9	Dimethyl hydrogen phosphite
756-79-6	Dimethyl methylphosphonate
124-64-1	Tetrakis(hydroxymethyl)phosphonium chloride
55566-30-8	Tetrakis(hydroxymethyl)phosphonium sulfate
1330-78-5	Tricresyl phosphate
512-56-1	Trimethyl phosphate

126-72-7	Tris(2,3-dibromopropyl) phosphate
78-51-3	Tris(2-butoxyethyl) phosphate
79-95-8	2,2',6,6'-Tetrachlorobisphenol A
118-79-6	2,4,6-Tribromophenol
56803-37-3	tert-Butylphenyl diphenyl phosphate
13674-84-5	Tris(2-chloroisopropyl)phosphate
25155-23-1	TXP
598-72-1	2-Bromopropionic acid
3194-55-6	1,2,5,6,9,10-Hexabromocyclododecane
6145-73-9	Tris(2-chloropropyl) phosphate
26040-51-7	Bis(2-ethylhexyl) tetrabromophthalate
68937-41-7	Phenol, isopropylated, phosphate (3:1)
2781-11-5	Phosphonic acid, [[bis(2-hydroxyethyl)amino]methyl]-, diethyl ester
78-30-8	Tri-o-cresyl phosphate
4162-45-2	Ethanol, 2,2'-((1-methylethylidene)bis((2,6-dibromo-4,1-phenylene)oxy))bis-

Are there enriched chemotypes within this Flame Retardant subset relative to ToxCast & Tox21?

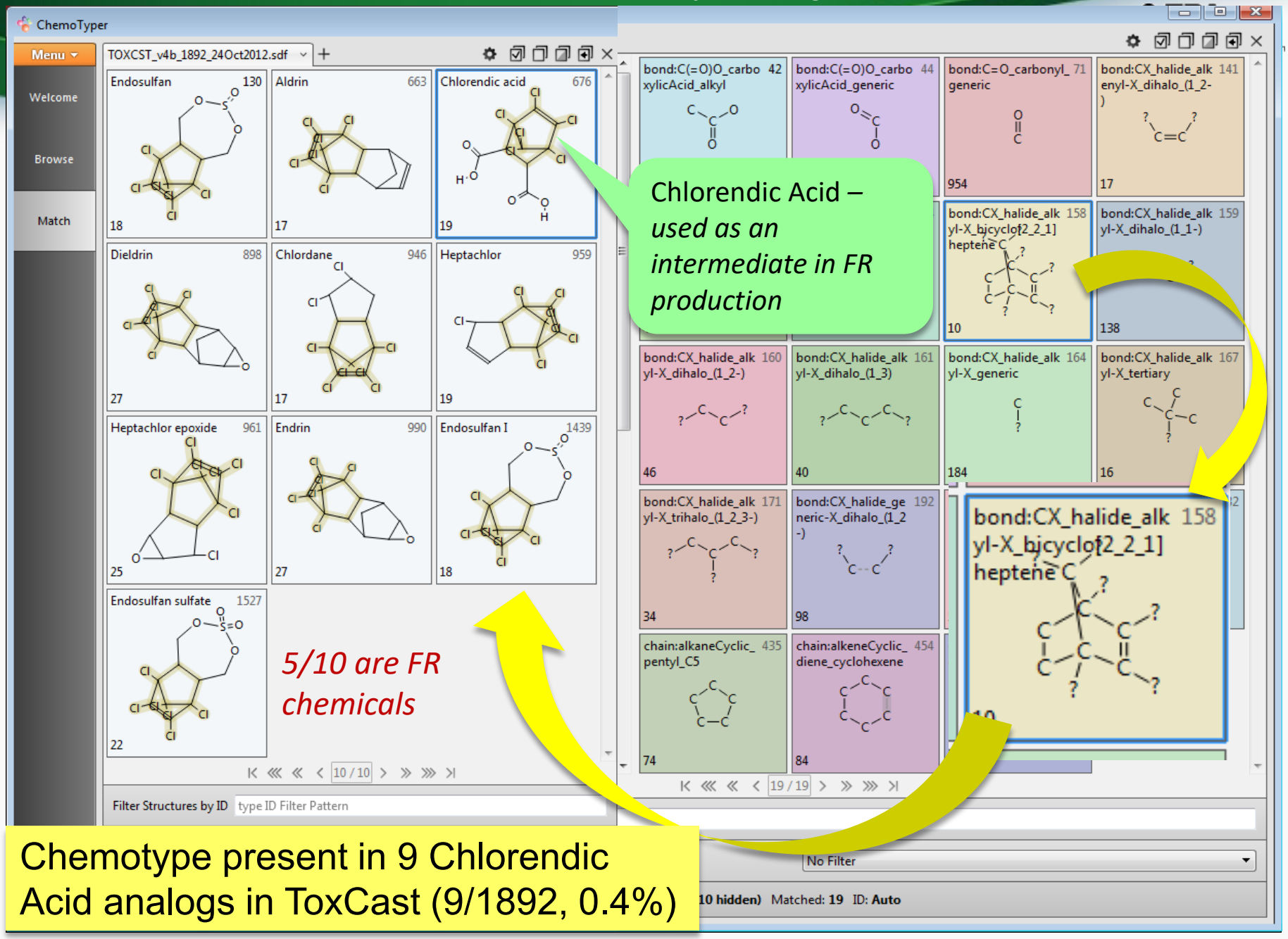
w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

Chemotype Enrichment,
e.g. *Flame Retardant (FR) Use Category*



w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

www.chemotyper.org



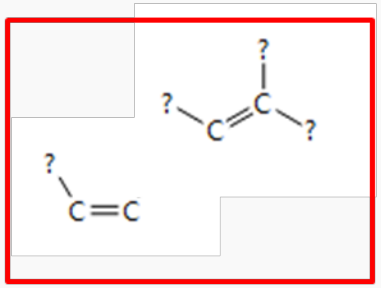
Chlorendic Acid – *used as an intermediate in FR production*

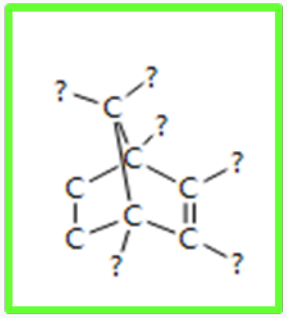
5/10 are FR chemicals

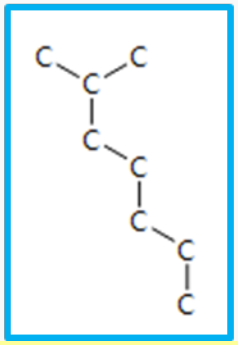
Chemotype present in 9 Chlorendic Acid analogs in ToxCast (9/1892, 0.4%)

{CT-Assay}$_{ToxCast\text{-}FR}$

Assay	FR Chemotype (CT)	Odds Ratio	# CT TP	# CT total	# assay positive	Total cmpds
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_2-)	6	8	46	71	1857
ATG_RARa_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_2-)	5	8	46	77	1857
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_3)	8	9	40	71	1857
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_trihalo_(1_2_3-)	9	8	34	71	1857
ATG_RARa_TRANS	bond:CX_halide_alkyl-X_trihalo_(1_2_3-)	5	6	34	77	1857
ATG_RXRb_TRANS	chain:alkaneBranch_isooctyl_hexyl_2-methyl	6	8	17	261	1857
Tox21_TR_LUC_GH3_Antagonist	bond:CX_halide_alkyl-X_bicyclo[2_2_1]heptene	18	6	10	151	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkenyl-X_dihalo_(1_2-)	10	12	17	385	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkyl-X_bicyclo[2_2_1]heptene	35	9	10	385	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkyl-X_tertiary	8	10	15	385	1858





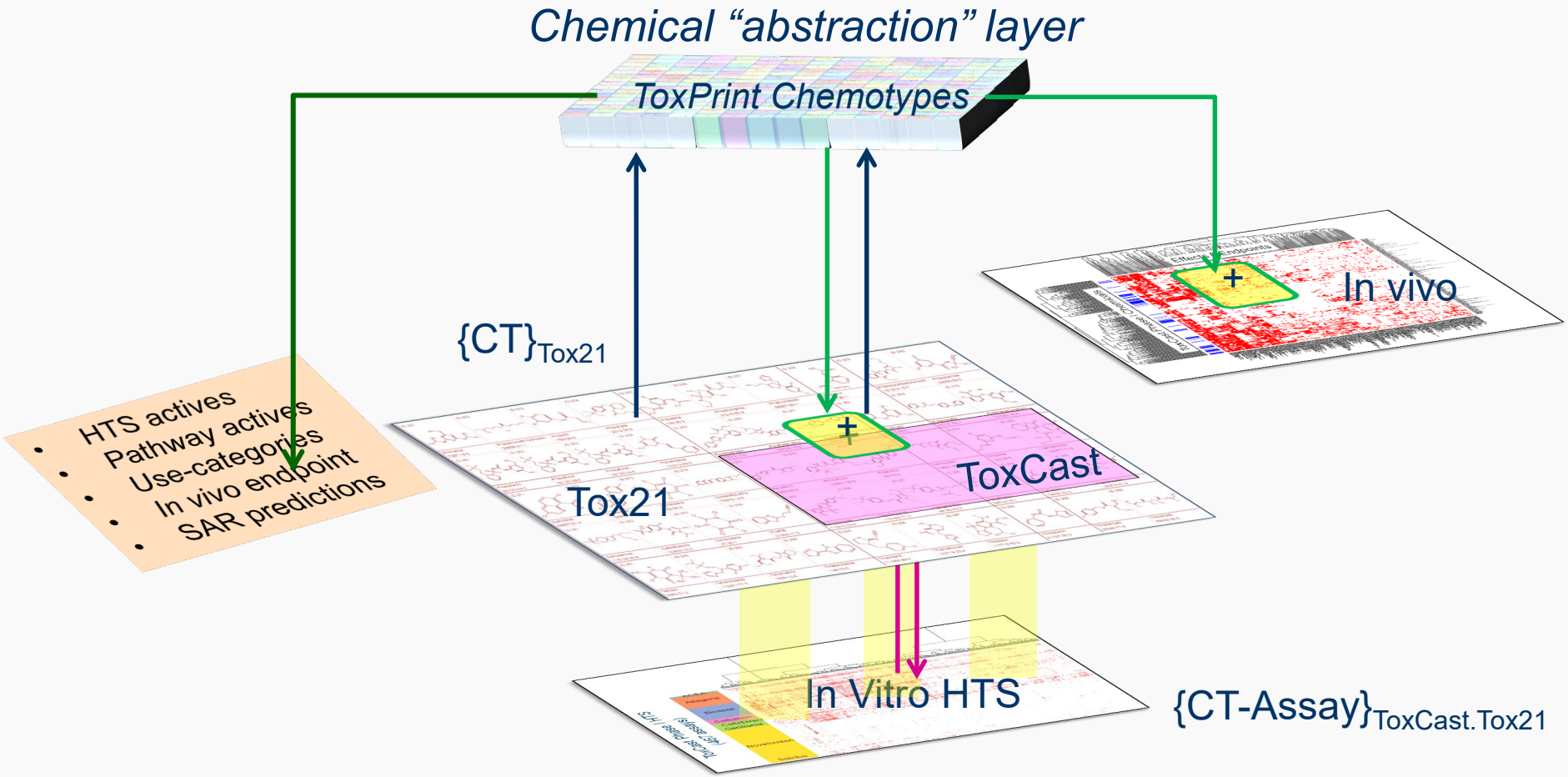


Significant {CT-Assay}$_{ToxCast\text{-}FR}$ associations potentially related to developmental outcomes

w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

Chemotype-Activity Enrichments

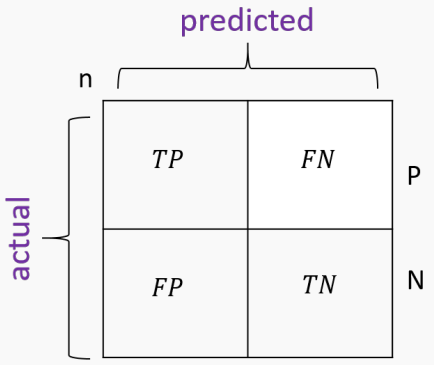
- ✓ Create {CT-Activity} enrichment profiles for any “active” subspace of a test set
- ✓ Focus studies in local CT domains & compare enrichments across data domains



Computing CT-Assay "Enrichments"

Set statistical thresholds & filters for significance to support data-mining objectives:

TP_ID	ToxPrint_CT_name[2]	CT_{Tot}	T_{pos}	F_{pos}	F_{neg}	T_{neg}	Odd's Ratio	Fischer's pval
423	chain:alkaneBranch_t-butyl_C4	41	24	17	294	693	3.3	2.0E-04
479	chain:aromaticAlkane_Ph-C1-Ph	39	27	12	291	698	5.4	6.5E-07
303	bond:X[any_!C]_halide_inorganic	28	17	11	301	699	3.6	9.0E-04



Confusion matrix

TestSet = # Pos + # Neg = # chems tested

CT_{Tot} = total # chems in TestSet w/ CT (Pos or Neg)
TP (T_{pos}) = # Pos in TestSet w/ CT
FP (F_{pos}) = # Neg in TestSet w/ CT
FN (F_{neg}) = # Pos in TestSet w/o CT
TN (T_{pos}) = # Neg in TestSet w/o CT

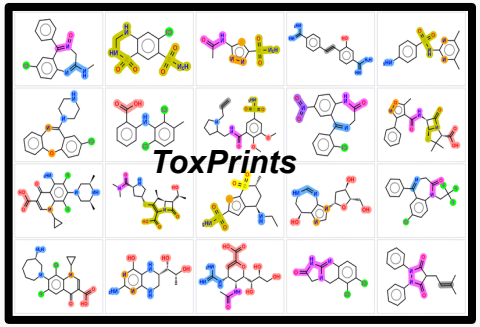
- Odds Ratio ≥ 3, *conveys simple fractional enrichment*
- Fischer's exact p value ≤ 0.05, *takes into account size of dataset*
- T_{pos} (TP) ≥ 3, *require at least 3 chemicals with CT in Positives*

Automated Chemotype-Enrichment Workflow (CTEW)

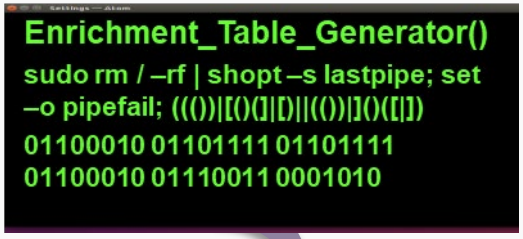
ID	Endpoint
DTXCID50675904	0
DTXCID50675924	0
DTXCID50675929	0
DTXCID50675904	1
DTXCID50675924	1
DTXCID50675929	0
DTXCID50675904	1
DTXCID50675924	0
DTXCID50675929	0
DTXCID50675904	0
DTXCID50675924	0
DTXCID50675929	0

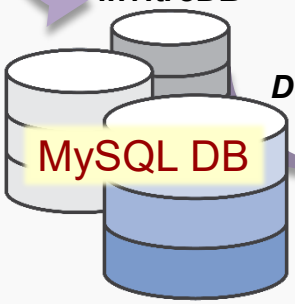
Generate binary "Hitcall Matrix"

Generate fingerprints



Calculate enrichment statistics





Store CTEW results in DB

Generate CT-Enrichment Table Results

descriptors_name	label	CT-Tot	TP	FP	FN	TN	BA	OR	P-Val	Inv OR	Inv P-val
Txp-1	atom:element_main_group	2	1	1	1827	5180	0.6196303963661194	2.8352489471435547	0.4536220133304596	0.35270270705223083	0.9320069551467896
Txp-10	bond:C#N_cyano_cyanohydrin	5	1	4	1827	5177	0.46957454085350037	0.7084017395973206	0.7794185876846313	1.4116283655166626	0.6100181937217712
Txp-100	bond:CN_amine_pri-NH2_alkyl	112	25	87	1803	5094	0.4808981418609619	0.8118652701377869	0.8468714356422424	1.2317314147949219	0.21217265725135803
Txp-101	bond:CN_amine_pri-NH2_aromatic	341	84	257	1744	4924	0.4923933148384094	0.9228215217590332	0.7525069713592529	1.0836331844329834	0.28962212800979614
Txp-102	bond:CN_amine_pri-NH2_generic	459	110	349	1718	4832	0.4886806607246399	0.8864842653274536	0.8697083592414856	1.1280516386032104	0.1554924845695495
Txp-103	bond:CN_amine_sec-NH_alkyl	199	44	155	1784	5026	0.4795689284801483	0.7997395992279053	0.9174827933311462	1.2504069805145264	0.11147873103618622
Txp-104	bond:CN_amine_sec-NH_aromatic	151	36	115	1792	5066	0.4885549545288086	0.8849766850471497	0.7643690705299377	1.129973292350769	0.298209011554718
Txp-105	bond:CN_amine_sec-NH_aromatic_aliphatic	97	22	75	1806	5106	0.4827597141265869	0.82932448387146	0.8107256293296814	1.2058006525039673	0.2608259320259094
Txp-106	bond:CN_amine_sec-NH_generic	253	58	195	1770	4986	0.48362982273101807	0.8378617763519287	0.8932034373283386	1.1935142278671265	0.13699662685394287
Txp-107	bond:CN_amine_ter-N_aliphatic	449	161	288	1667	4893	0.5522294044494629	1.6408655643463135	0.00000143775969263515 44	0.6094344258308411	0.9999991655349731
Txp-108	bond:CN_amine_ter-N_aromatic	167	66	101	1762	5080	0.568841278553009	1.8839976787567139	0.00008360712672583759	0.5307862162590027	0.9999573826789856

Slide courtesy of Ryan Lougee

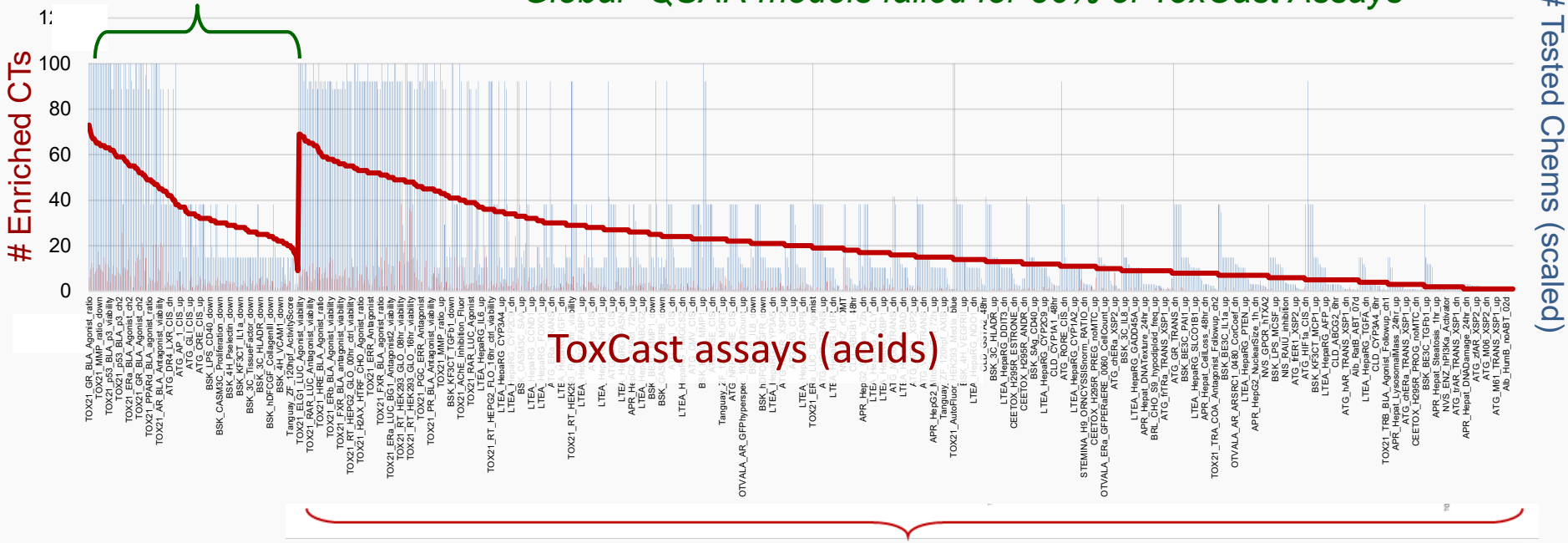
Assays x Enriched CT count

Enriched CT-counts across 1032 assays

*Random Forest models based on ToxPrint CT descriptors, validated using independent Test Set & Y-randomization, with Training (100A,100I) & Test (25A,25I) Set minimums (J. Fitzpatrick)

"Global" QSAR models* for 167 ToxCast Assays
Avg % Median BA = 0.67

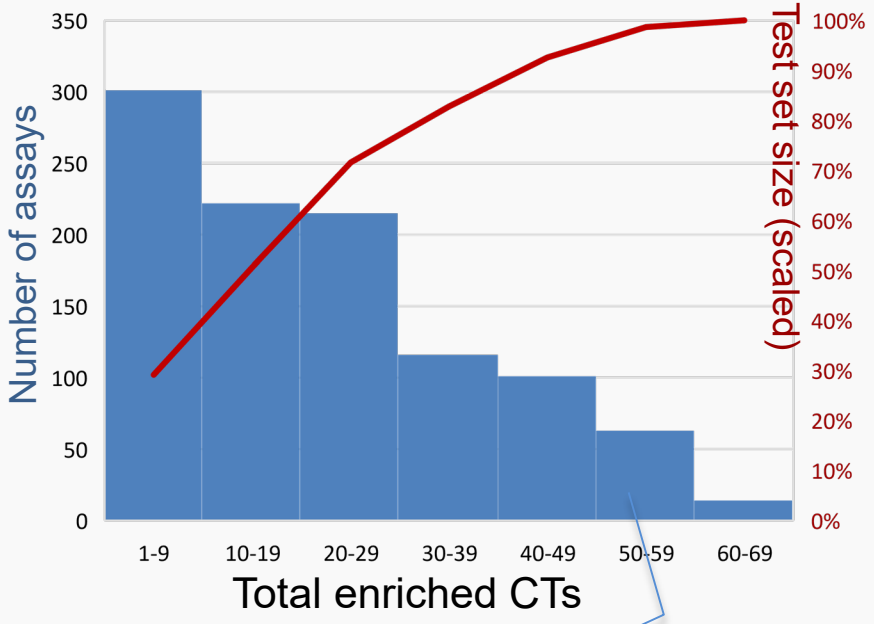
"Global" QSAR models failed for 86% of ToxCast Assays



Significant CT-enrichments across ToxCast assay space considered "unmodelable" by traditional "global" QSAR methods

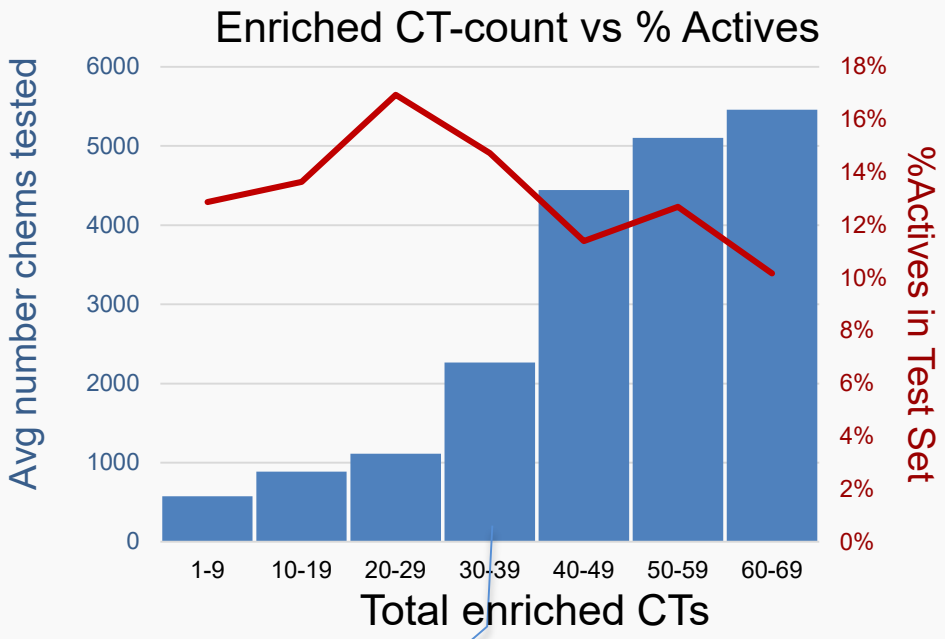
Enriched CT-count per assay trends (1032 assays)

Does # enriched CTs depend on test set size? YES!



63 assays enriched with 50-59 CTs

Does # enriched CTs depend on %Actives? Not so much.



30-39 CTs enriched in assays with avg Test Set size 2264 chems

Tuning biology to pick up chemical "signal"

EPA United States Environmental Protection Agency

CT Enrichments of Time-series* "Assays"

aenm_870unique	TP_Count	%actives	tested.chnm
APR_Hepat_Apoptosis_1hr_up	1	3.2	310
APR_Hepat_Apoptosis_24hr_up	15	14.8	310
APR_Hepat_Apoptosis_48hr_up	6	16.8	310
APR_Hepat_DNADamage_1hr_up	8	3.9	310
APR_Hepat_DNADamage_24hr_up	7	18.4	310
APR_Hepat_DNADamage_48hr_up	14	20.0	310

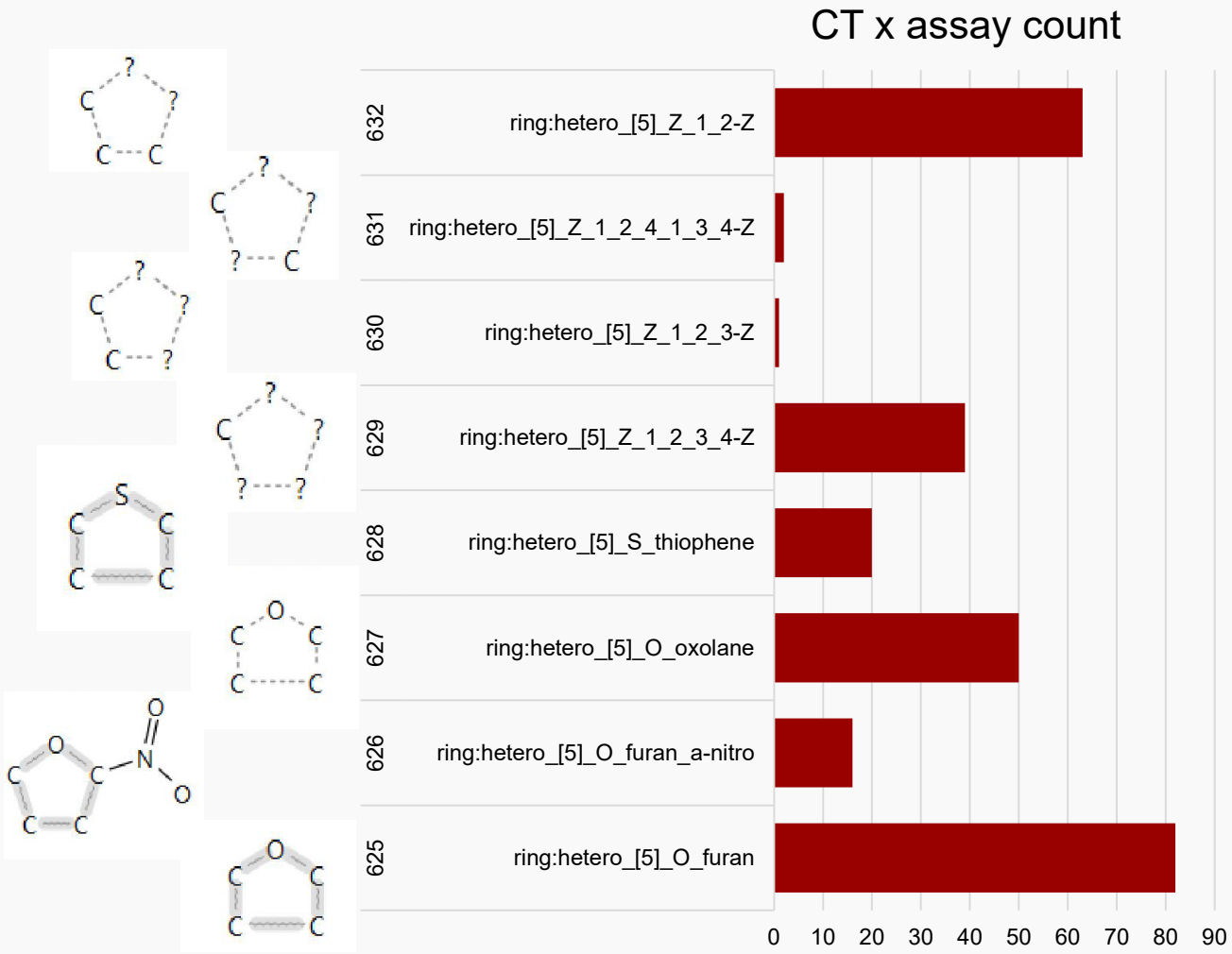
** 95 Assays out of 1032 total are within time-series groups*

"Tuning" biology to increase CT enrichments → increases biological-chemical "signal"

- How are {CT-assay} enrichments affected by cell toxicity ("burst") filters?
- How are {CT-assay} enrichments affected by activity threshold assumptions?

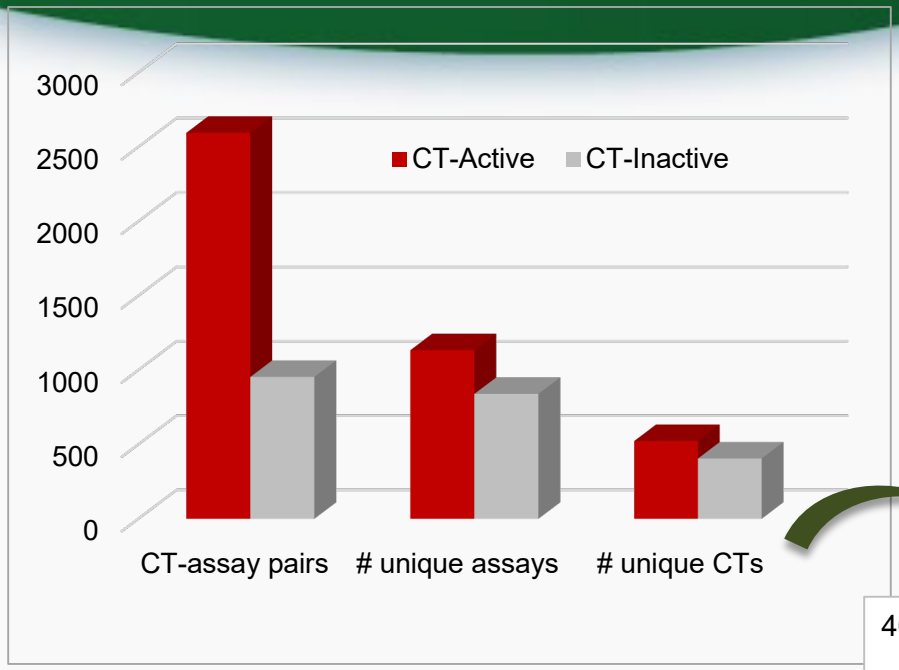
Tuning chemistry to pick up biological "signal"

Assay Enrichments of chemistry-series groups



"Tuning" chemistry specificity to increase chemical-assay "signal"

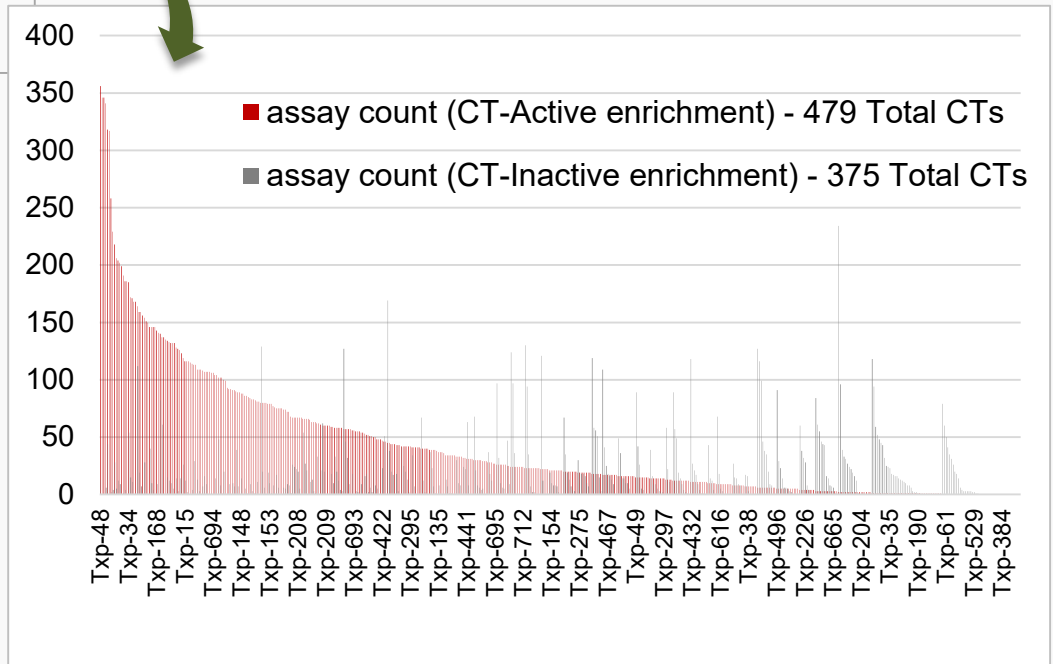
What about the "Inactives"?



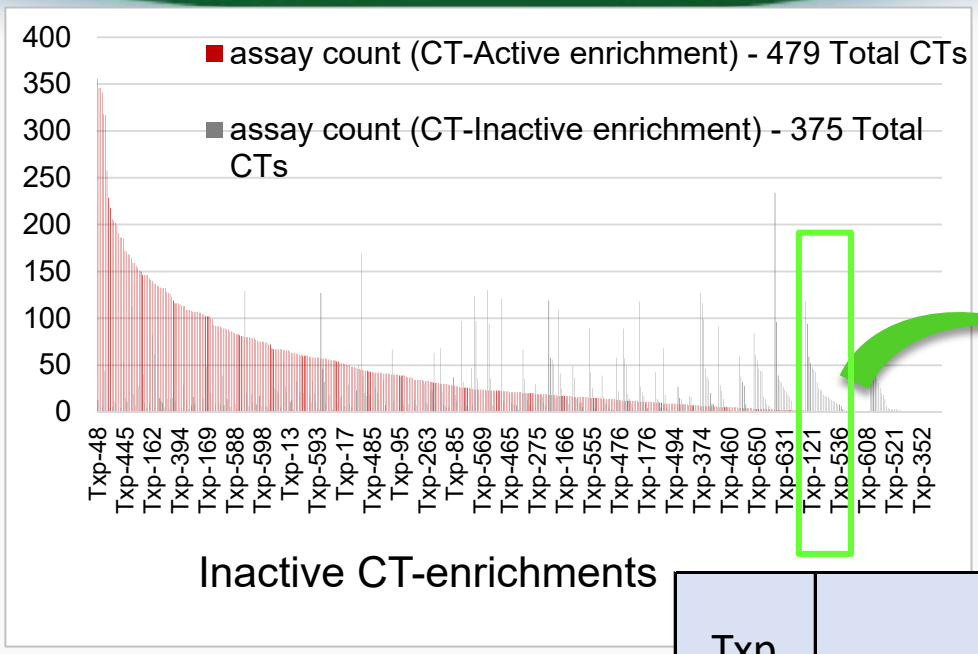
More "CT-signal" in Actives, but significant amount of signal in Inactives

Inactive CT-enrichments

- Span ToxCast assay space
- More frequently occur in assays with fewer Active CT-enrichments
- May be due to several factors:
  - *True inactivity*
  - *Assay artifacts*
  - *QC failure*



Top 10 enriched CT-Inactives (skewed from actives)

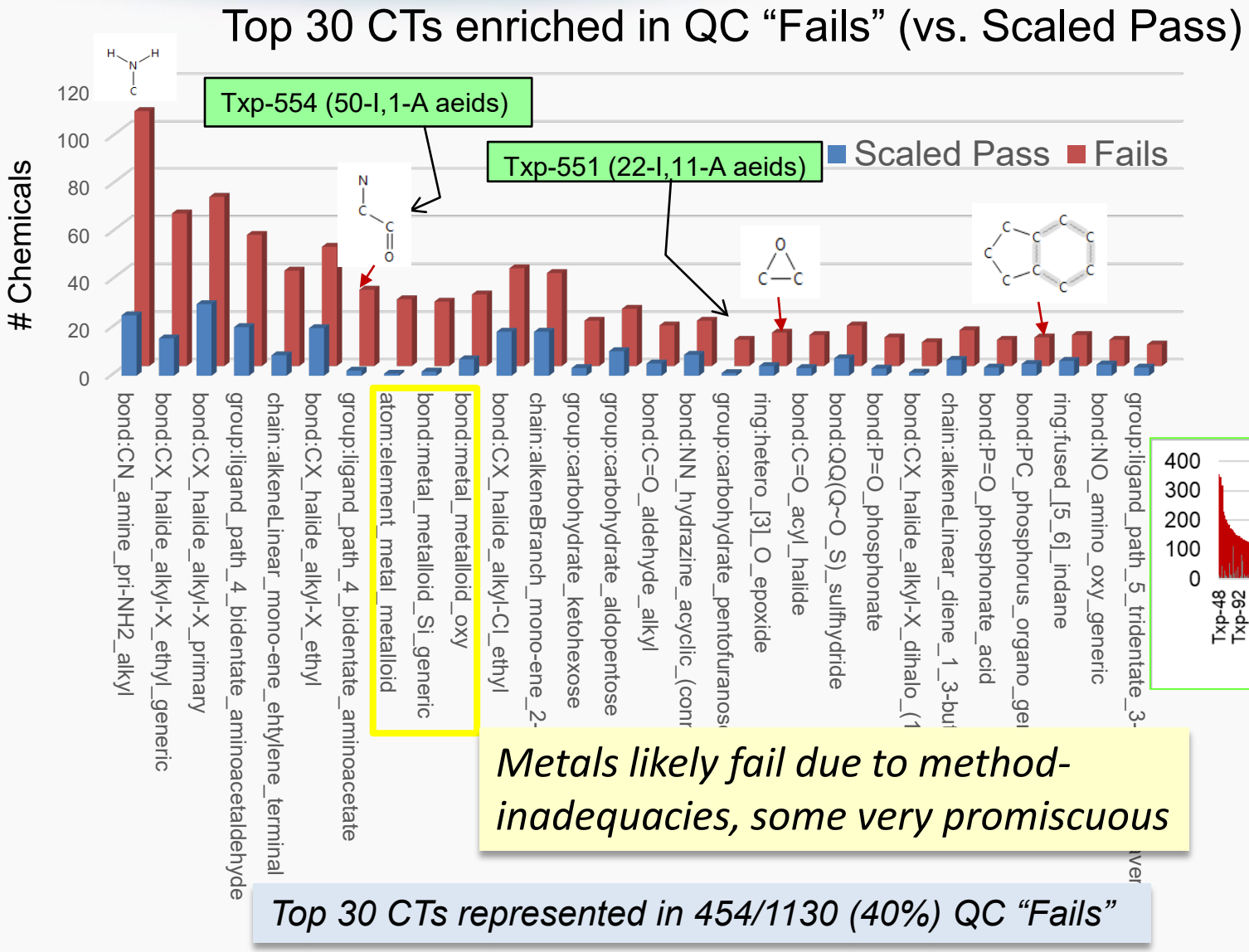


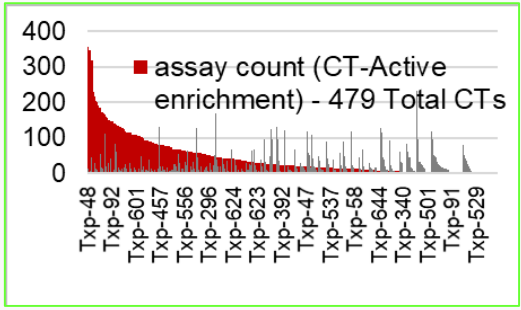
CT-Inactive >75 assays, CT-active <10 assays

- *True inactivity?*
- *Assay artifacts?*
- ***QC failure?***

Txp	CT Name	assay count (CT-Active enrichment)	assay count (CT-Inactive enrichment)
Txp-101	bond:CN_amine_pri-NH2_aromatic	3	84
Txp-145	bond:CX_halide_alkyl-Cl_ethyl	0	79
Txp-260	bond:P~S_generic	1	94
Txp-362	bond:metal_metalloid_oxy	2	96
Txp-372	bond:metal_metalloid_Si_organo	6	99
Txp-374	bond:metal_metalloid_Si _oxy	6	116
Txp-496	chain:oxy-alkaneLinear_ethyleneOxide_EO1(O)	5	91
Txp-497	chain:oxy-alkaneLinear_ethyleneOxide_EO2	6	127
Txp-607	ring:hetero_[4]_N_beta_lactam	1	118
Txp-663	ring:hetero_[6]_N_triazine_(1_3_5-)	2	234

ToxPrints enriched in Fails





Metals likely fail due to method-inadequacies, some very promiscuous

Top 30 CTs represented in 454/1130 (40%) QC "Fails"

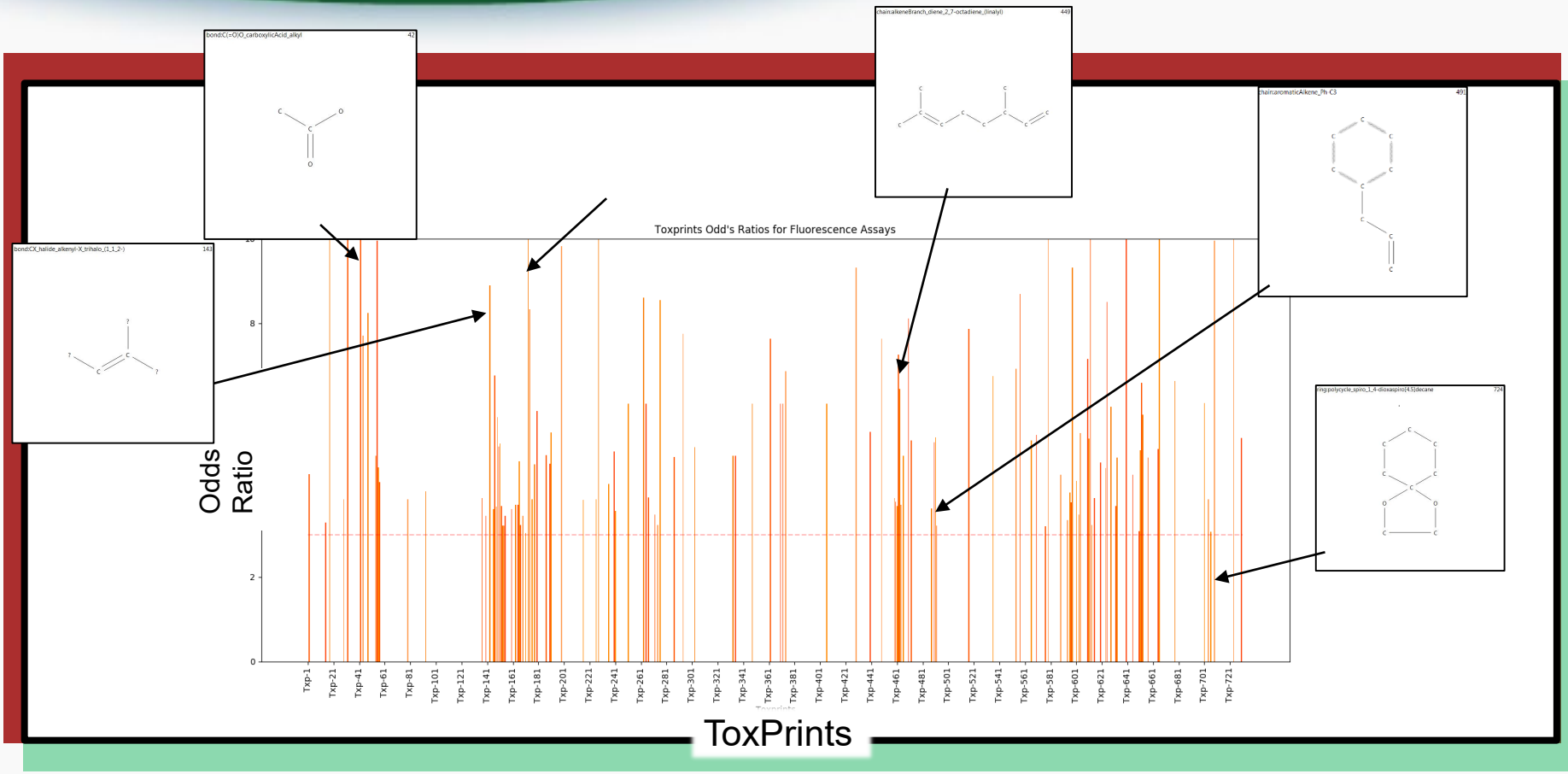
Detection Technology Type: Fluorescence

Fluorescence: Detection techniques that use the principles of fluorescence, whereby incident light excites a fluorophore which then emits light at lower energy (higher wavelength). The emitted light is typically from the visible portion of the UV-Visible spectrum.

Table 5: ToxCast detection technology types.
Values are counts of assay endpoints

Detection technology type	subtype	Totals
Fluorescence	Fluorescence intensity	601
	FRET: TR-FRET	8
	Fluorescence polarization	1
Radiometry	Scintillation counting	136
Spectrophotometry	Absorbance	40
Luminescence	Bioluminescence	28
	Chemiluminescence	1
Microscopy	Optical microscopy: Fluorescence microscopy	4
Label-Free Technology	Electrical Sensor: Impedance	2

CT Enrichments in Fluorescence Assays



CT's enriched (much more likely to be active) in "Fluorescence Assays" than in remaining assays

#7:

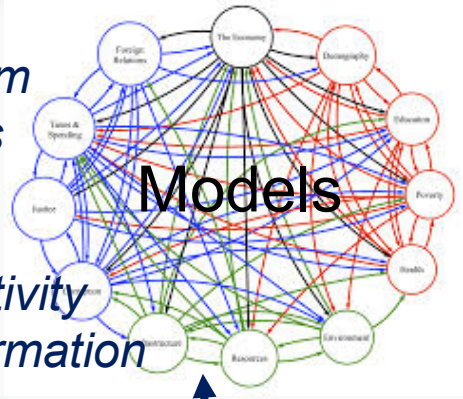
Until we understand the limits of our data, we can't make best use of it.

Building a public chemotype "knowledge- base"

EPA United States Environmental Protection Agency

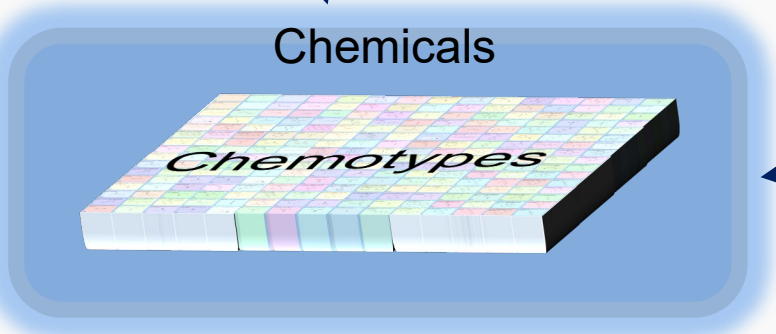
Phys-chem properties

Exposure Data



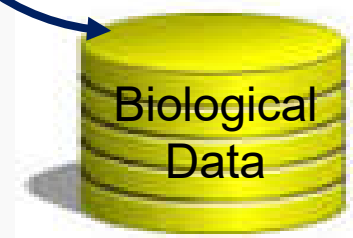
Reactivity
Biotransformation

Fate & Transport
ADME



Toxicity Predictions
Structure-alerts

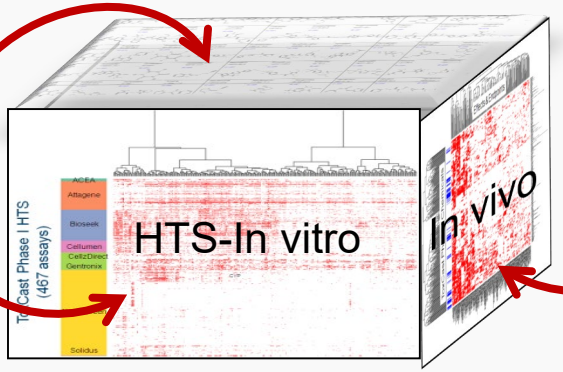
Biological activities



CPCat, Use Exposure

Chemicals

ToxCast
Tox21



ToxRef
COSMOS

EPA's CompTox Chemicals Dashboard

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

Journal of Cheminformatics

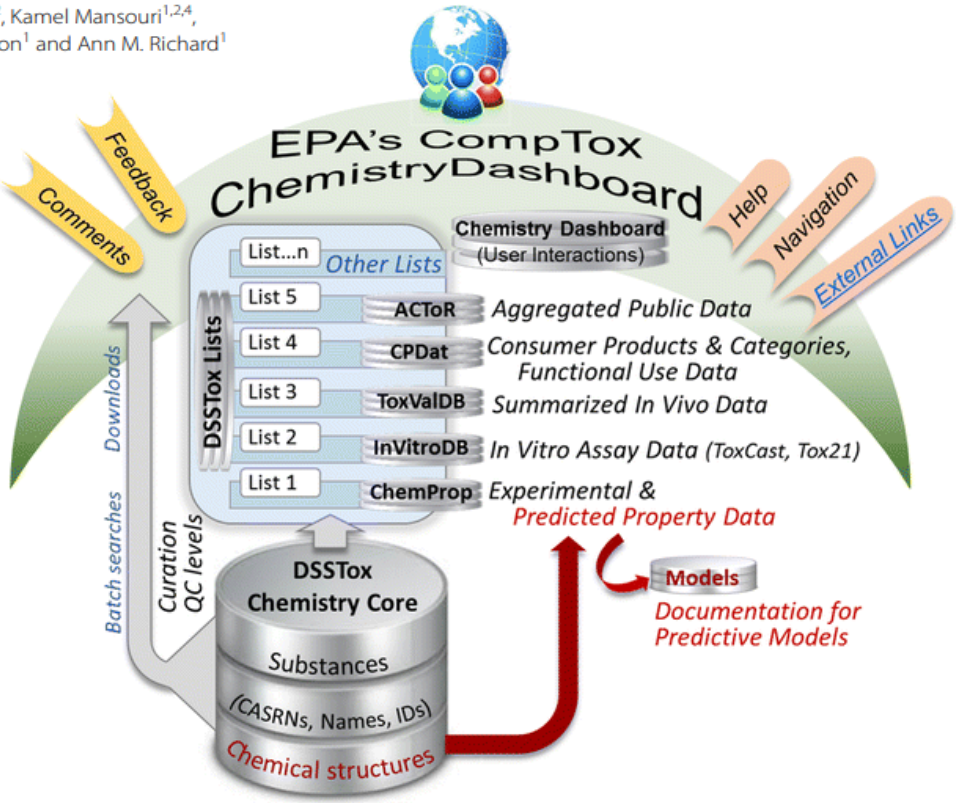
DATABASE Open Access

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

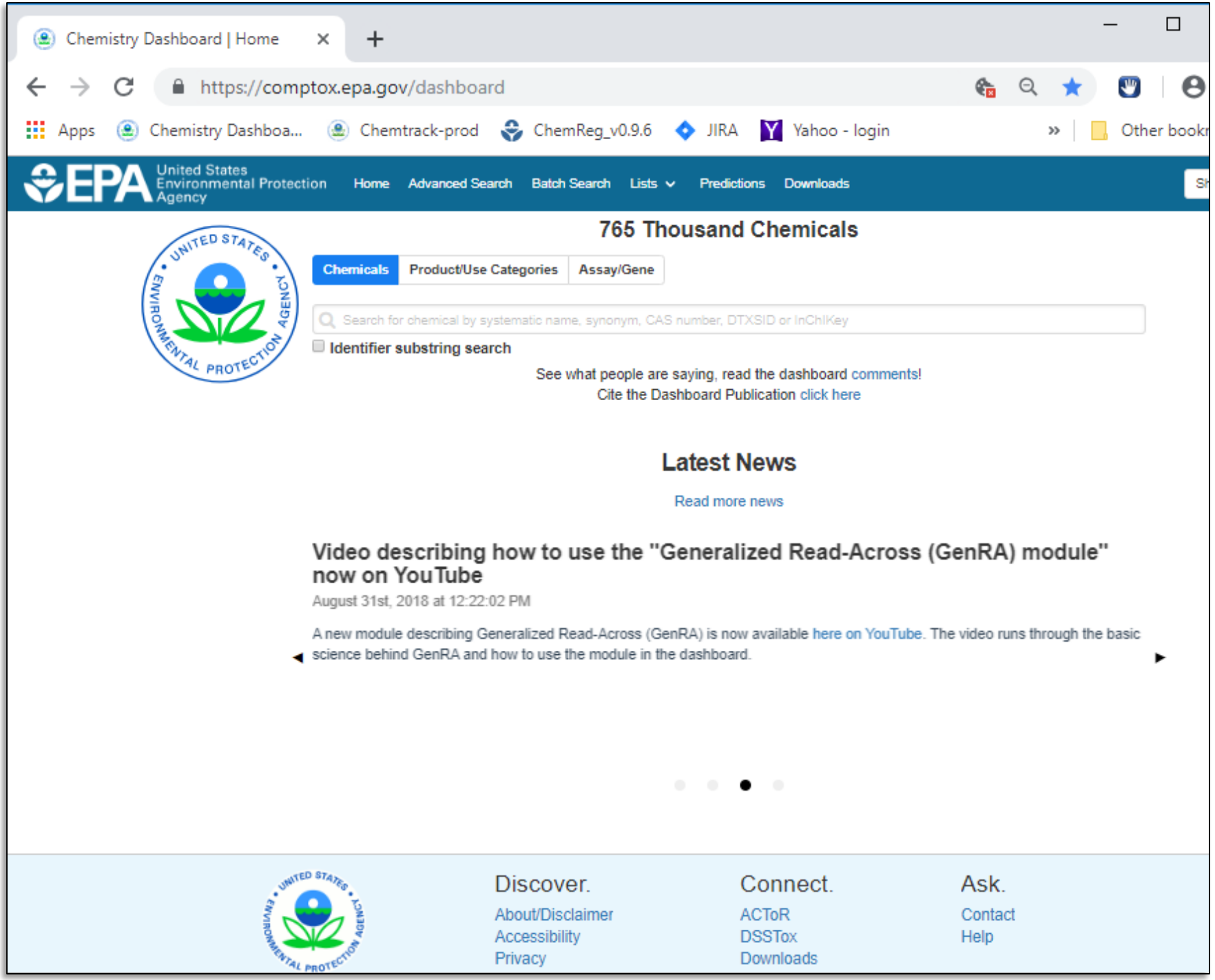
Antony J. Williams[1*], Christopher M. Grulke[1], Jeff Edwards[1], Andrew D. McEachran[2], Kamel Mansouri[1,2,4], Nancy C. Baker[3], Grace Patlewicz[1], Imran Shah[1], John F. Wambaugh[1], Richard S. Judson[1] and Ann M. Richard[1]

Open Access free download at:

https://jcheminf.biomedcentral.com/articles/10.1186/s13321-017-0247-6



https://doi.org/10.1186/s13321-017-0247-6
*J. Chem. Inf. 2017, **9**: 61-88.*


Chemistry Dashboard | Home
https://comptox.epa.gov/dashboard
Apps
Chemistry Dashboa...
Chemtrack-prod
ChemReg_v0.9.6
JIRA
Yahoo - login
Other bookr
United States Environmental Protection Agency
Home
Advanced Search
Batch Search
Lists
Predictions
Downloads
765 Thousand Chemicals
Chemicals
Product/Use Categories
Assay/Gene
Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey
Identifier substring search
See what people are saying, read the dashboard comments!
Cite the Dashboard Publication click here
Latest News
Read more news
Video describing how to use the "Generalized Read-Across (GenRA) module" now on YouTube
August 31st, 2018 at 12:22:02 PM
A new module describing Generalized Read-Across (GenRA) is now available here on YouTube. The video runs through the basic science behind GenRA and how to use the module in the dashboard.
Discover.
About/Disclaimer
Accessibility
Privacy
Connect.
ACToR
DSSTox
Downloads
Ask.
Contact
Help

Chemistry Dashboard

https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID9020112

Apps | Chemistry Dashboa... | Chemtrack-prod | ChemReg_v0.9.6 | JIRA | Yahoo - login | Evernote Sign-in | Dropbox | Other bookmarks

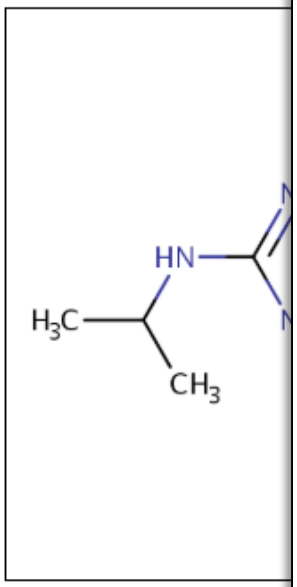
EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search

Atrazine

1912-24-9 | DTXS

Searched by DSSTox Substan



- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS

Presence in Lists

Federal

- ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
- Endocrine Disruptor Screening Program (EDSP) Universe of Chemicals
- TSCA Inventory, active non-confidential portion
- EDSP21 Tier 1 Screening Chemicals: List 1
- ECOTOXicology knowledgebase (ECOTOX)
- TOXCAST - EPA ToxCast Screening Library
- CHEMINV: ToxCast/Tox21 Chemical inventory available as DMSO solutions (20181123)
- ATSDR Toxic Substances Portal Chemical List
- Superfund Chemical Data Matrix
- NIOSH International Chemical Safety Cards
- EPA Integrated Risk Information System (IRIS)
- Toxics Release Inventory
- NIOSH Pocket Guide to Chemical Hazards
- TOXCAST_PhaseI - EPA ToxCast Screening Library (Phase I subset)
- TOXCAST_PhaseII - EPA ToxCast Screening Library (Phase II Subset)
- TOXCAST_PhaseIII - EPA ToxCast Screening Library (Phase II Subset)
- TOXCAST_ph1v2 - EPA ToxCast Screening Library (ph1v2 Subset)
- EPA Pesticide Chemical Search Database

US State

- California Office of Environmenta

International

- SUSDAT: The NORMAN Network

Other

- Androgen Receptor Chemicals
- Wiley Registry of Tandem Mass Spectral Data, MSforID
- Thermo's mzCloud Database
- CERAPP: Collaborative Estrogen Receptor Activity Prediction Project
- Integrated Biological Pathway model for the Estrogen Receptor
- CPDAT, the EPA Chemical and Products Database
- Hydrogen Deuterium Exchange Standard Set - No Exchange
- TOX21SL: Tox21 Screening Library
- NormaNEWS: Norman Early Warning System
- University Jaume I Target Substances
- MassBank Reference Spectra Collection
- NORMAN Collaborative Trial 2015 Targets and Suspects
- STOFF-IDENT Database of Water-Relevant Substances
- MassBank.EU Collection: Special Cases
- National Environmental Methods Index
- University of Athens Target List
- EPA's Drinking Water Standard and Health Advisories Table

Chemistry Dashboard

https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID9020112#related-substances

Apps | Chemistry Dashboa... | Chemtrack-prod | ChemReg_v0.9.6 | JIRA | Yahoo - login | Evernote Sign-in | Dropbox | CompToxDashboar... | Other bookmarks

Atrazine

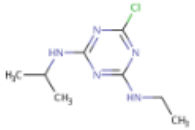
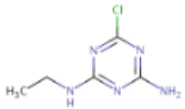
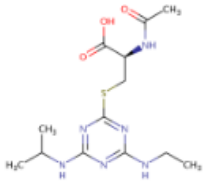
1912-24-9 | DTXSID9020112

Searched by DSSTox Substance Id.

4 chemicals

Download / Send | Show info: DTXSID | CASRN | TOXCAST | Select all

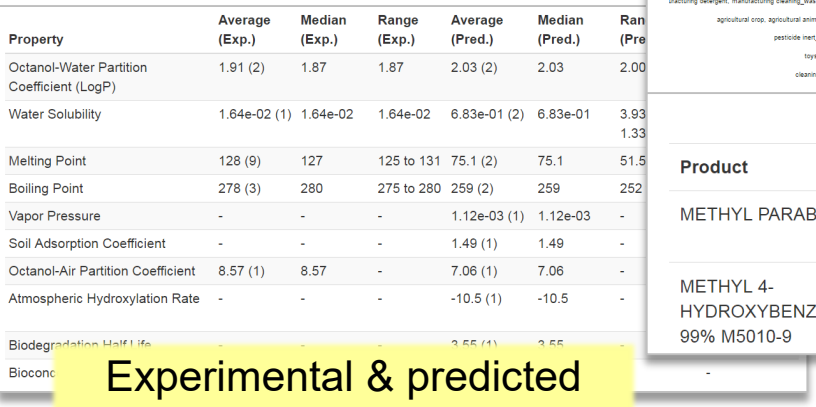
Sort by: Relationship | Filter by: Name or CASRN | Hide

Searched Chemical	Predecessor: Component	Transformation Product	Transformation Product
	2 related chemical structures with this substance		
Atrazine DTXSID: DTXSID9020112 CASRN: 1912-24-9 TOXCAST: 32/729	Propachlor - Atrazine mixture DTXSID: DTXSID70896770 CASRN: 8070-76-6 TOXCAST: 0	Deisopropylatrazine DTXSID: DTXSID0037495 CASRN: 1007-28-9 TOXCAST: 19/713	Atrazine mercapturate DTXSID: DTXSID90160763 CASRN: 138722-96-0 TOXCAST: 0

EPA's CompTox Dashboard

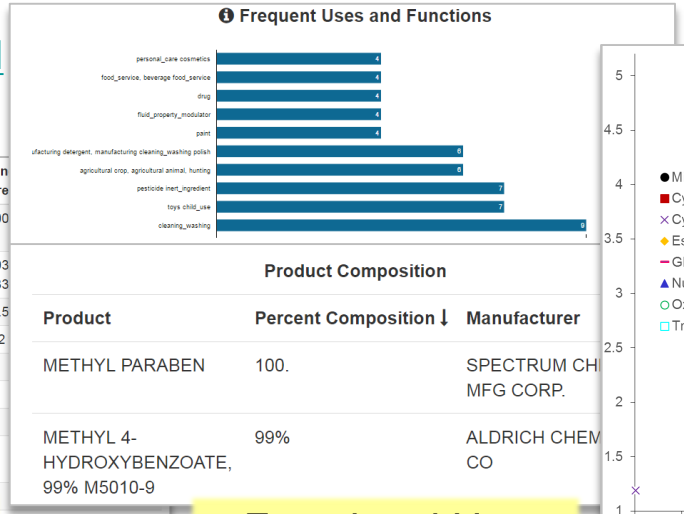
williams.antony@epa.gov

https://comptox.epa.gov/dashboard

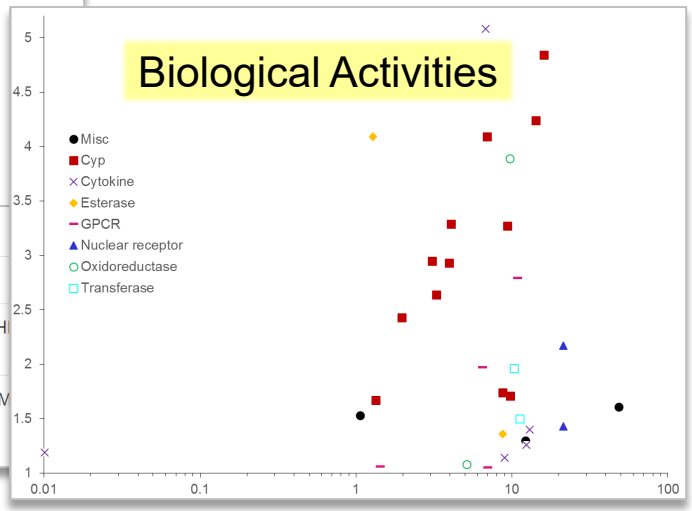


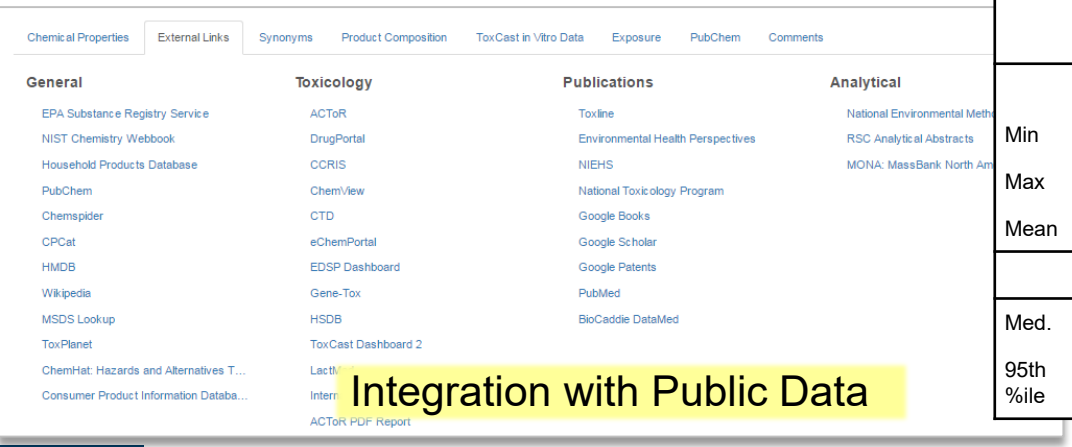
Property	Average (Exp.)	Median (Exp.)	Range (Exp.)	Average (Pred.)	Median (Pred.)	Ran (Pre
Octanol-Water Partition Coefficient (LogP)	1.91 (2)	1.87	1.87	2.03 (2)	2.03	2.00
Water Solubility	1.64e-02 (1)	1.64e-02	1.64e-02	6.83e-01 (2)	6.83e-01	3.93 1.33
Melting Point	128 (9)	127	125 to 131	75.1 (2)	75.1	51.5
Boiling Point	278 (3)	280	275 to 280	259 (2)	259	252
Vapor Pressure	-	-	-	1.12e-03 (1)	1.12e-03	-
Soil Adsorption Coefficient	-	-	-	1.49 (1)	1.49	-
Octanol-Air Partition Coefficient	8.57 (1)	8.57	-	7.06 (1)	7.06	-
Atmospheric Hydroxylation Rate	-	-	-	-10.5 (1)	-10.5	-
Biodegradation Half Life	-	-	-	2.55 (1)	2.55	-
Bioconc						-

Experimental & predicted phys-chem properties



Functional Use & Composition





Integration with Public Data

National Health and Nutrition Examination Survey (NHANES) Inferences (mg/kg-bw/day)							
	Ages 6-11	Ages 65+	BMI > 30	BMI < 30	Females	Males	Total
Min	6.82e-04	5.12e-04	6.57e-04	9.81e-04	1.51e-03	4.64e-04	8.95e-04
Max	1.01e-03	7.95e-04	8.41e-04	1.14e-03	1.89e-03	5.89e-04	1.01e-03
Mean	8.28e-04	6.36e-04	7.44e-04	1.06e-03	1.68e-03	5.25e-04	9.53e-04
Exposure Predictions (mg/kg-bw/day)							
Med.	1.88e-05	1.35e-05	1.46e-05	2.05e-05	2.37e-05	1.55e-05	1.98e-05
95th %ile	2.00e-03	1.25e-03	1.30e-03	1.63e-03	1.44e-03	1.14e-03	1.09e-03

Exposure

#8:

Making quality data, structures, and capabilities freely available is the best way to win friends and influence people.

#9:
What a difference the right people at the right time can make.

The Past …

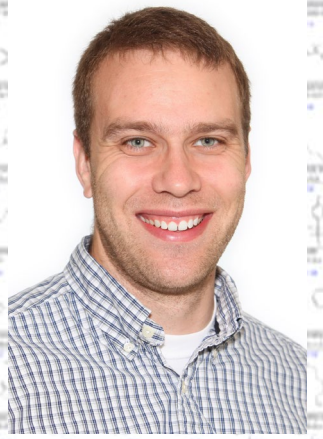
ClarLynda Williams
DSSTox_v0

Jamie Burch
DSSTox website

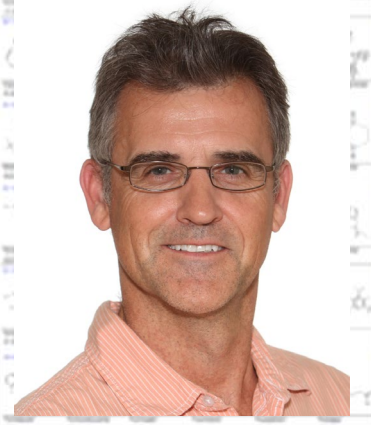
Dr. Marty Wolf
(1947-2012)
DSSTox_v1
ToxCasts, Tox21

"Ain't going in the database unless it's done right"

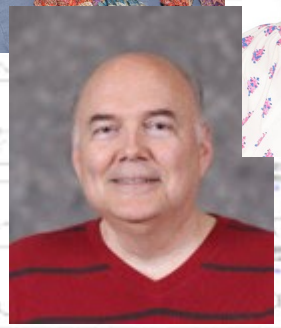
The Present …



Chris Grulke
Cheminformatics Genius

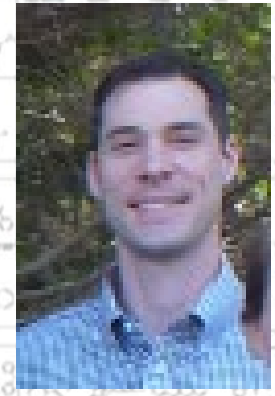


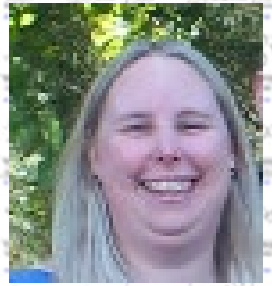
Tony Williams
Dashboard Hero



Indira Thillainadarajah
Saku Sivasupramaniam
Brian Meyer
DSSTox Super-Curators

Ryan Lougee
Chemotype-Enrichment Workflow Master





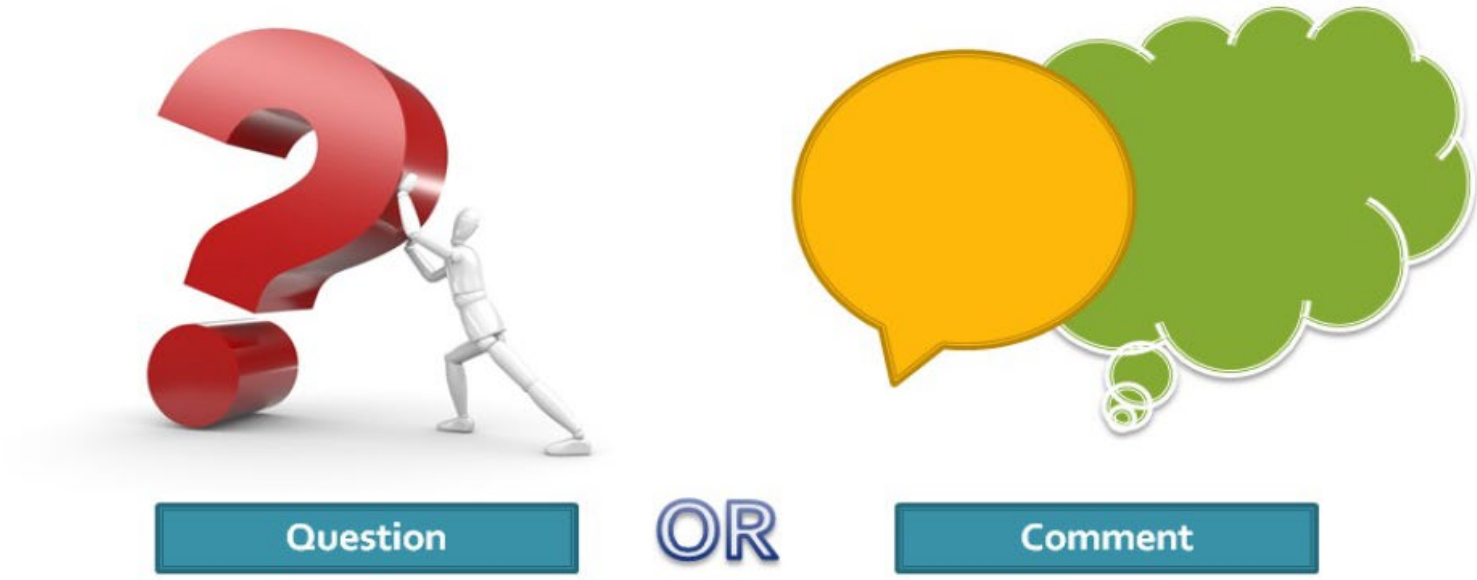
Jon Sobus & Elin Ulrich
Non-Targeted Screening (ENTACT) Partners in Crime

EPA's National Center for Computational Toxicology
Research Triangle Park, NC

NCCT's ToxCast Team

EPA
United States
Environmental Protection
Agency

Thank you for your attention



EPA Training Opportunities for recent BA graduates & Post Docs

https://orise.orau.gov/epa/applicants/current-research-opportunities.aspx