United States Environmental Protection Agency

US EPA Computational Toxicology Programs: Central Role of Chemical-annotation Efforts & Molecular Databases

ACS Fall Meeting, Washington DC, August 17, 2009

Ann Richard
richard.ann@epa.gov

Office of Research and Development
National Center for Computational Toxicology

EPA
United States
Environmental Protection
Agency

"…to integrate modern computing and information technology with molecular biology to improve Agency prioritization of data requirements and risk assessment of chemicals"

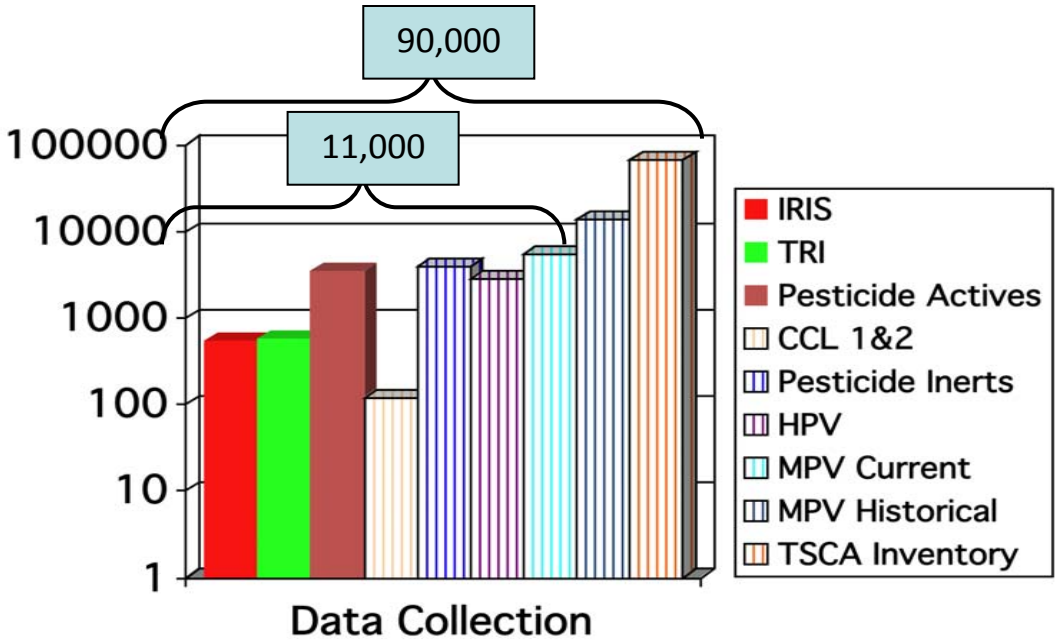
Decision Support Tools for High-Throughput Risk Assessment

Office of Research and Development
National Center for Computational Toxicology

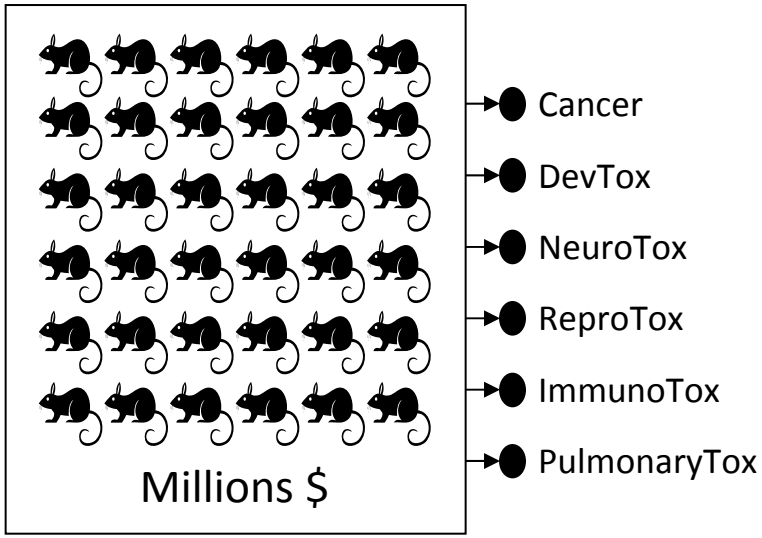
http://www.epa.gov/ncct

Change Needed Because …..

Too Many Chemicals



Too High a Cost



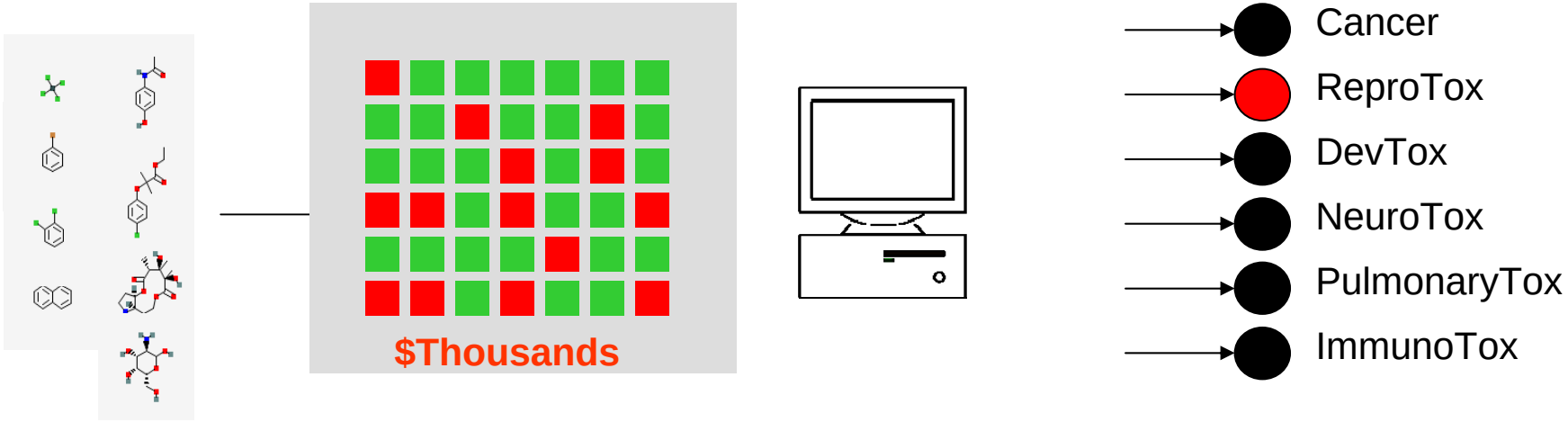
…and not enough data.

Judson, et al *EHP, 2008*

Future of Chemical Toxicity Testing

United States Environmental Protection Agency

in vitro testing *in silico* analysis



HTS
-omics

Bioinformatics/
Machine Learning

Future of Chemical Toxicity Testing

United States Environmental Protection Ag

THE NATIONAL

REP

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 Academy of Sciences • National Academy of Engineering • Institute of Medicine • National Research

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.

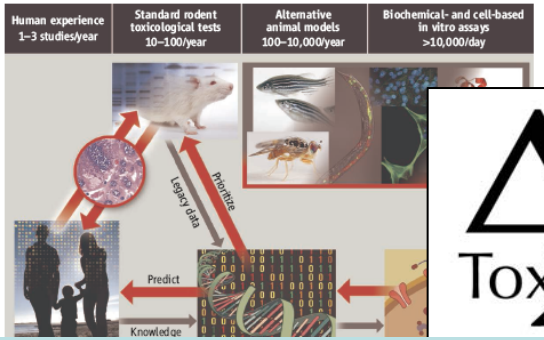
Francis S. Collins,[1*†] George M. Gray,[2*] John R. Bucher[3*]

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,



Tox21

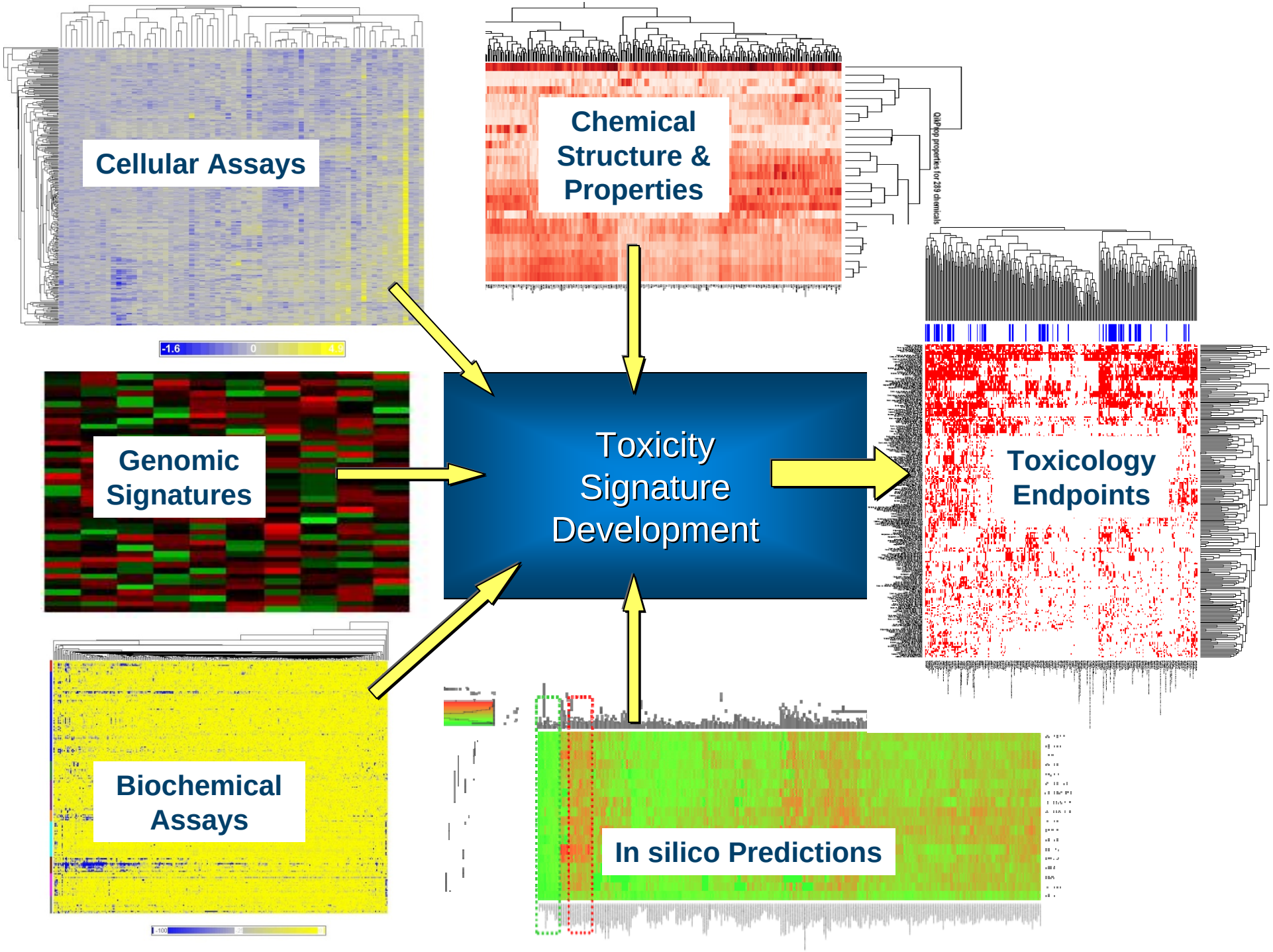
EPAs Contribution: The ToxCast Research Program

*The views expressed here are those of the individual authors and do not necessarily reflect the views and policies of their respective agencies.

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

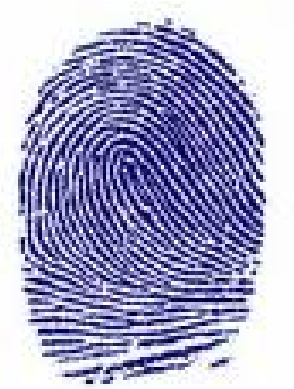
Transforming toxicology. The studies we propose will test whether high-throughput and computational toxicology approaches can yield data predictive of results from animal toxicity studies, will allow prioritization of chemicals for further testing, and can assist in prediction of risk to humans.

National Center for Computational Toxicology


Cellular Assays
Chemical Structure & Properties
QikProp properties for 289 chemicals
Genomic Signatures
Toxicity Signature Development
Toxicology Endpoints
Biochemical Assays
In silico Predictions

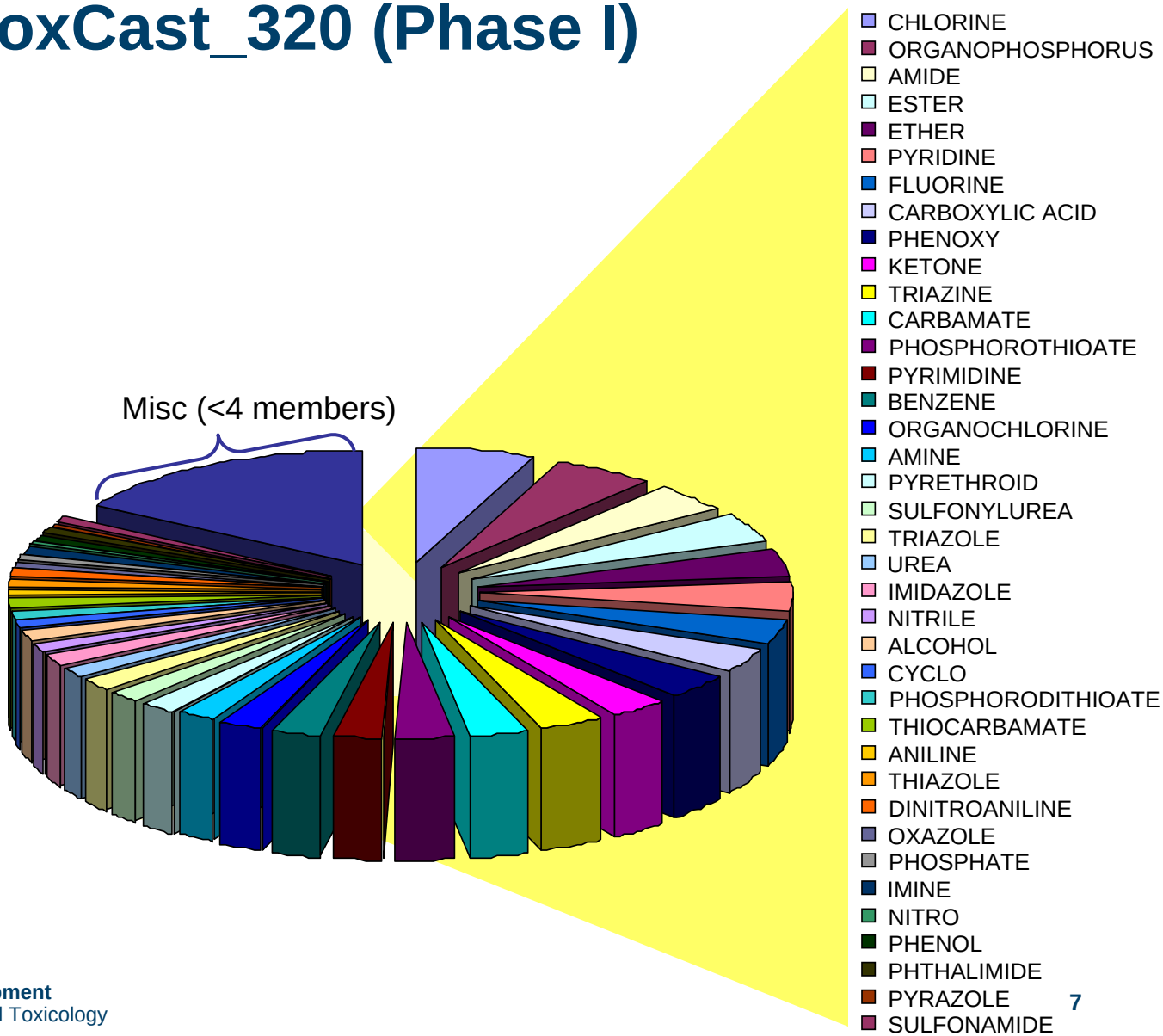
ToxCast™ Background

- Research program of EPA's National Center for Computational Toxicology (NCCT)
- Addresses chemical screening and prioritization needs for pesticidal inerts, anti-microbials, CCLs, HPVs and MPVs
- Comprehensive use of HTS technologies to generate biological fingerprints and predictive signatures
- Coordinated with NTP and NHGRI/NCGC via Tox21
- Committed to stakeholder involvement and public release of data
  - Communities of Practice- Chemical Prioritization; Exposure
  - NCCT website http://www.epa.gov/ncct/toxcast
    - *ACToR* *http://www.epa.gov/actor/*
    - *ToxRef DB* *http://www.epa.gov/ncct/toxrefdb/*
    - *DSSTox (PubChem)* *http://www.epa.gov/ncct/dsstox/*



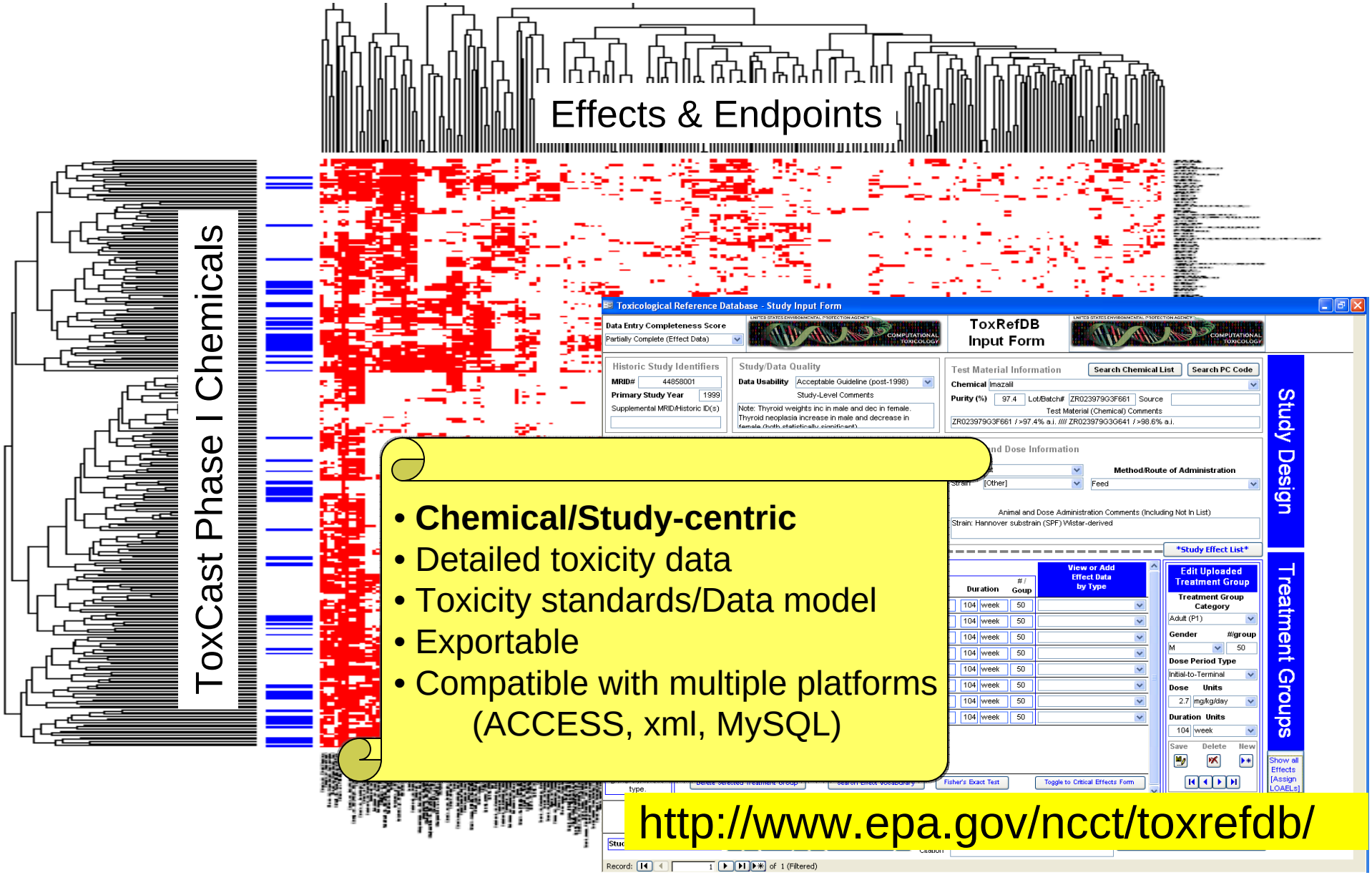
Chemical Classes in ToxCast_320 (Phase I)

- 309 Unique Structures
- Replicates for QC
- 291 Pesticide Actives
- 9 Industrial Chemicals
- 13 Parent/Metablolite pairs
- 56/73 Proposed Tier 1 Endocrine Disruption Screening Program
- 14 High Production Volume Chemicals
- 11 HPV Challenge



ToxRefDB: >$1Billion Million Dollars Worth of *In Vivo* Chronic/Cancer Bioassay Effects and Endpoints

United States Environmental Protection Agency



ToxCast *In vitro* HTS Assays

United States Environmental Protection Agency

Biochemical Assays

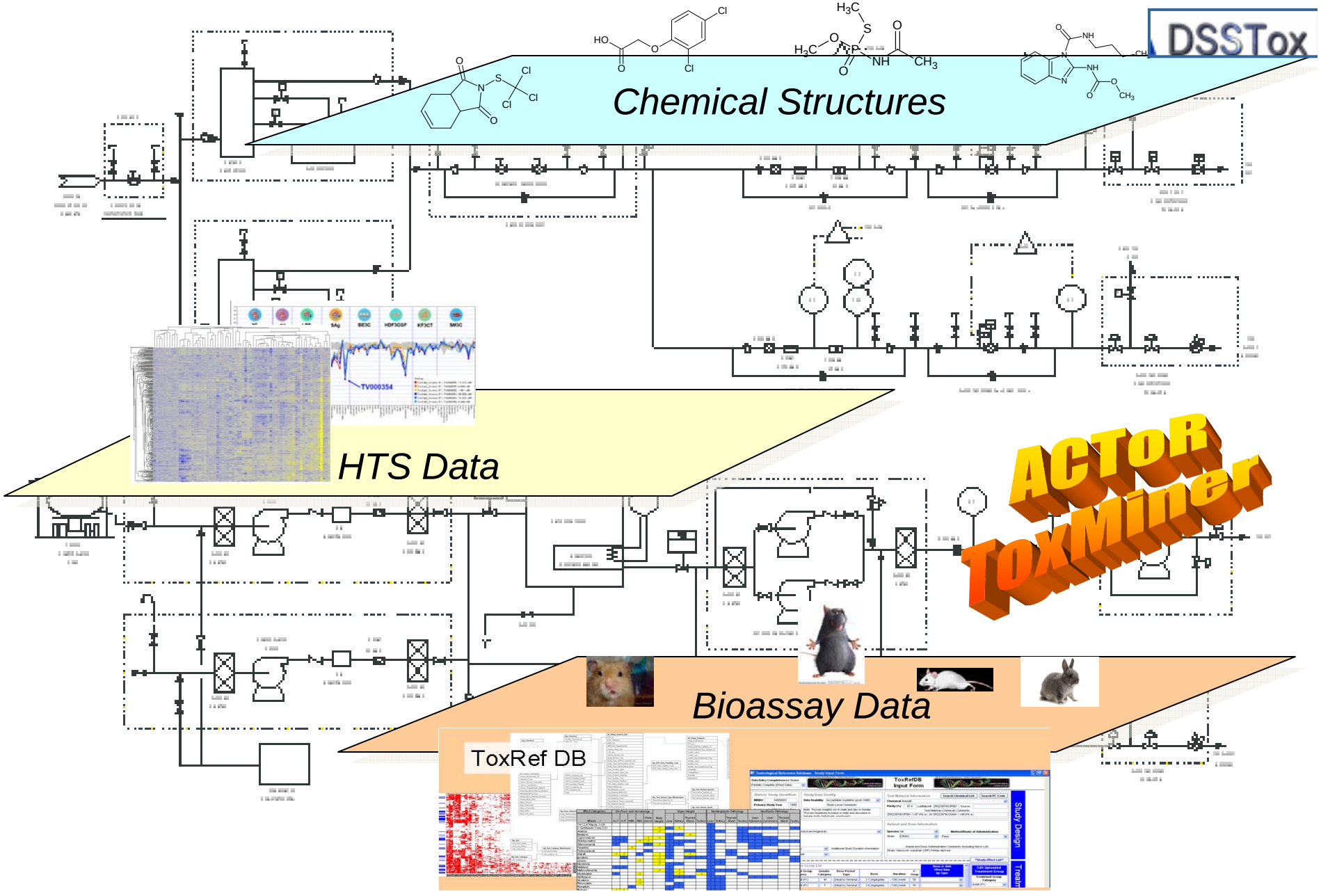
- **Protein families**
  - GPCR
  - NR
  - Kinase
  - Phosphatase
  - Protease
  - Other enzyme
  - Ion channel
  - Transporter
- **Assay formats**
  - Radioligand binding
  - Enzyme activity
  - Co-activator recruitment

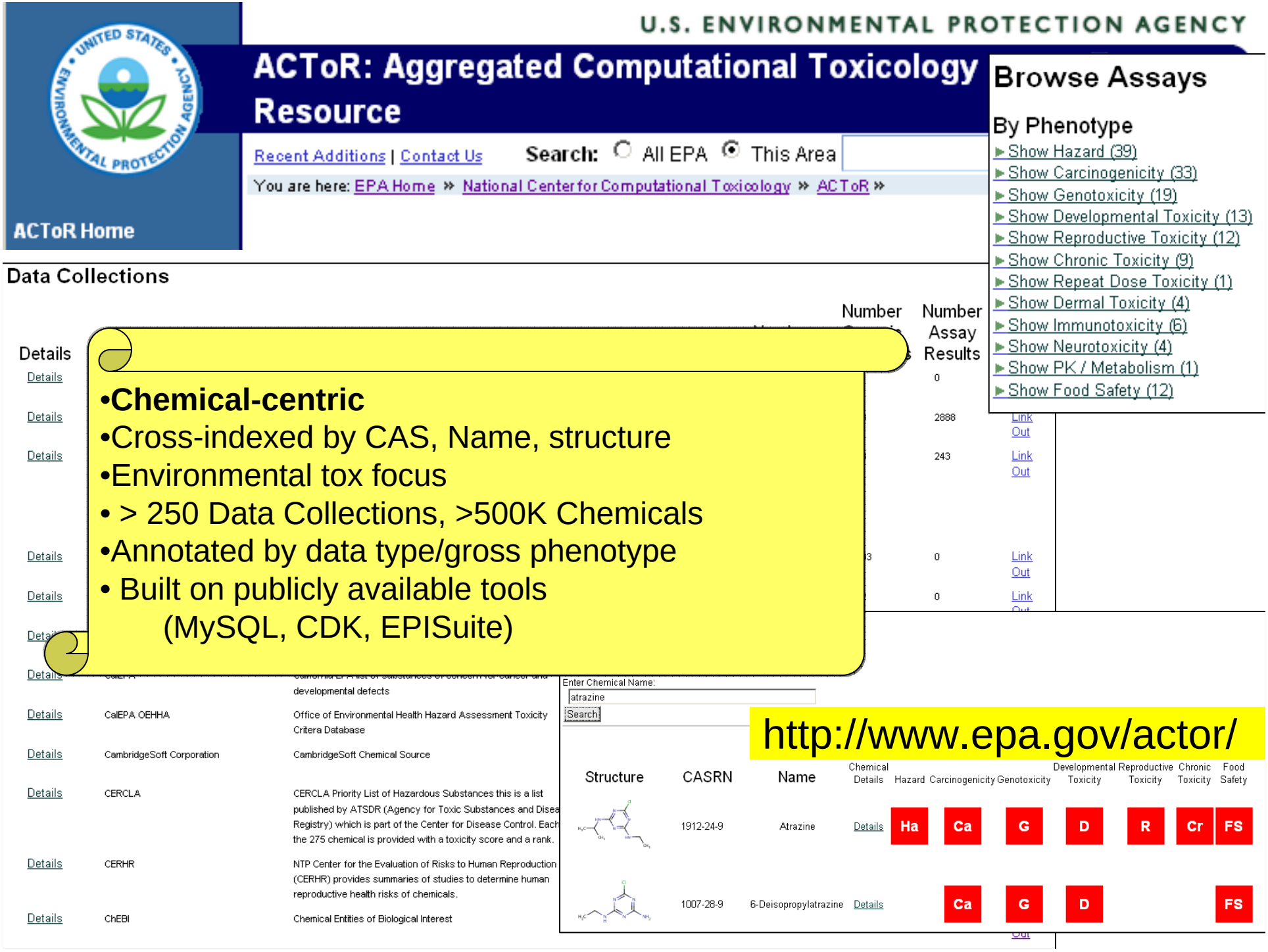
467 Endpoints

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
- **Biotransformation competent cells**
  - Primary rat hepatocytes
  - Primary human hepatocytes
- **Assay formats**
  - Cytotoxicity
  - Reporter gene
  - Gene expression
  - Biomarker production
  - High-content imaging for cellular phenotype

ToxCast: High-Multi-Dimensional Data



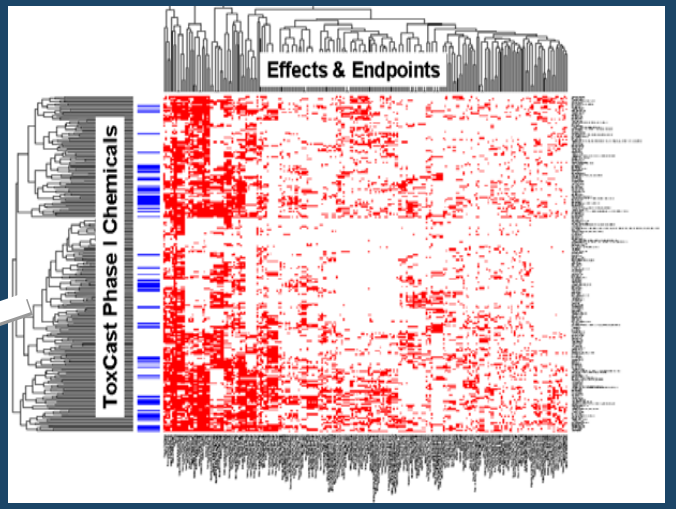

U.S. ENVIRONMENTAL PROTECTION AGENCY
ACToR: Aggregated Computational Toxicology Resource
ACToR Home
Recent Additions | Contact Us
Search: All EPA This Area
You are here: EPA Home » National Center for Computational Toxicology » ACToR »
Browse Assays
By Phenotype
Show Hazard (39)
Show Carcinogenicity (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
Number
Number Assay Results
•Chemical-centric
•Cross-indexed by CAS, Name, structure
•Environmental tox focus
• > 250 Data Collections, >500K Chemicals
•Annotated by data type/gross phenotype
• Built on publicly available tools
(MySQL, CDK, EPISuite)
developmental defects
CalEPA OEHHA
Office of Environmental Health Hazard Assessment Toxicity Critera Database
CambridgeSoft Corporation
CambridgeSoft Chemical Source
CERCLA
CERCLA Priority List of Hazardous Substances this is a list published by ATSDR (Agency for Toxic Substances and Disea Registry) which is part of the Center for Disease Control. Each the 275 chemical is provided with a toxicity score and a rank.
CERHR
NTP Center for the Evaluation of Risks to Human Reproduction (CERHR) provides summaries of studies to determine human reproductive health risks of chemicals.
ChEBI
Chemical Entities of Biological Interest
Enter Chemical Name:
atrazine
Search
http://www.epa.gov/actor/
Structure
CASRN
Name
Chemical Details
Hazard
Carcinogenicity
Genotoxicity
Developmental Toxicity
Reproductive Toxicity
Chronic Toxicity
Food Safety
1912-24-9
Atrazine
Details
Ha
Ca
G
D
R
Cr
FS
1007-28-9
6-Deisopropylatrazine
Details
Ca
G
D
FS

ToxCast & Tox21

ToxCast: Phase I

320 chemicals x
>450 HTS assays

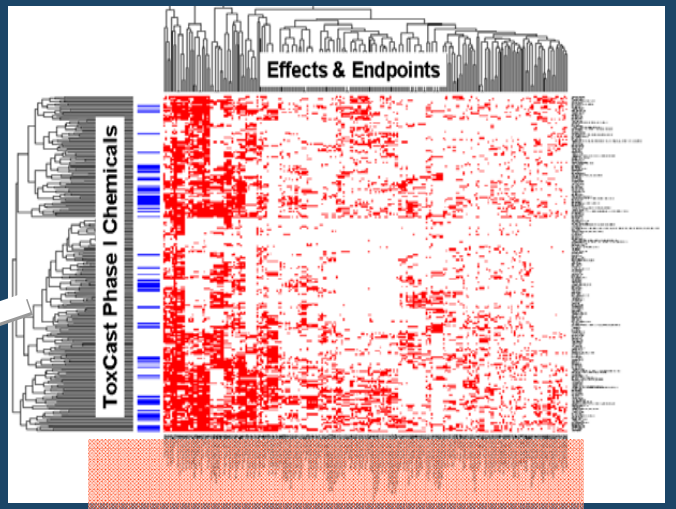
predictive signatures



ToxCast & Tox21

ToxCast: Phase I

320 chemicals x >450 HTS assays



predictive signatures

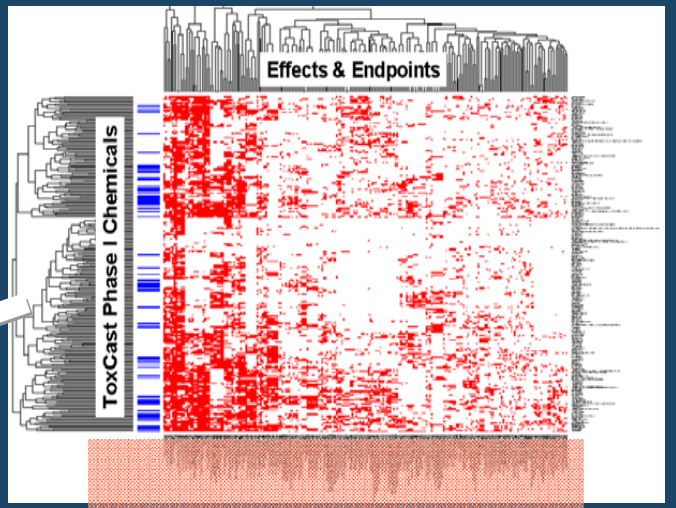
ToxCast: Phase II

~1000 chemicals >400 HTS assays

+ Public Tox Data

Refined predictive signatures

ToxCast & Tox21



ToxCast: Phase I

ToxCast: Phase II

320 chemicals x >450 HTS assays

predictive signatures

~1000 chemicals >400 HTS assays

+ Public Tox Data

Tox21

EPA

NTP

NCGC

Refined predictive signatures

~10,000 chemicals, 100s -1000s of HTS assays

Sparse Public Tox Data

ToxCast & Tox21

Chemical Sample Annotation & QC

Cast: ase I

ToxCast: Phase II

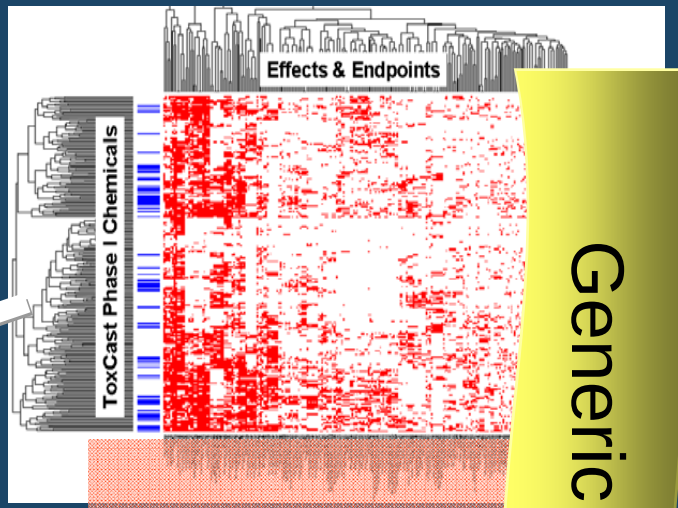
Tox21

EPA

NTP

NCGC

320 chemicals x >450 HTS assays



predictive signatures

~1000 chemicals >400 HTS assays

+ Public Tox Data

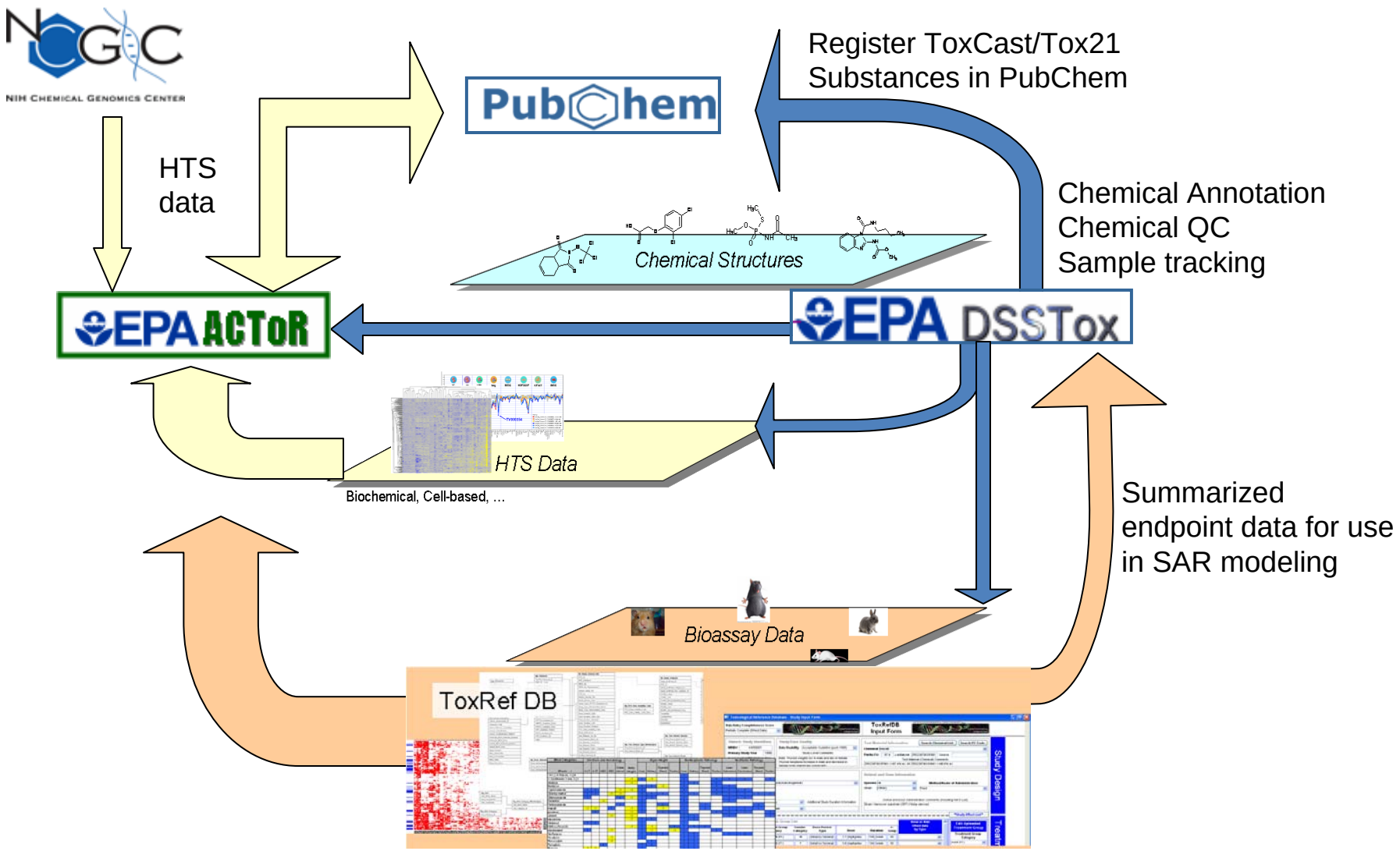
Refined predictive signatures

~10,000 chemicals, 100s -1000s of HTS assays

Sparse Public Tox Data

Generic Chemical Annotation

ToxCast/Tox21: Chemical Annotation & Data Publication



ToxCast/Tox21: Chemical Annotation & Data Publication

Cheminformatics Needs:

- Select chemicals for testing
  - ✓ high environmental / tox interest
  - ✓ suitable for testing
  - ✓ chemically diverse
  - ✓ analogs, metabolites, etc
- Standardized chemical structure/substance annotation across all testing & reference inventories
- Chemical annotation QC
- Analytical QC
- Public reporting of data & QC

otation

ng

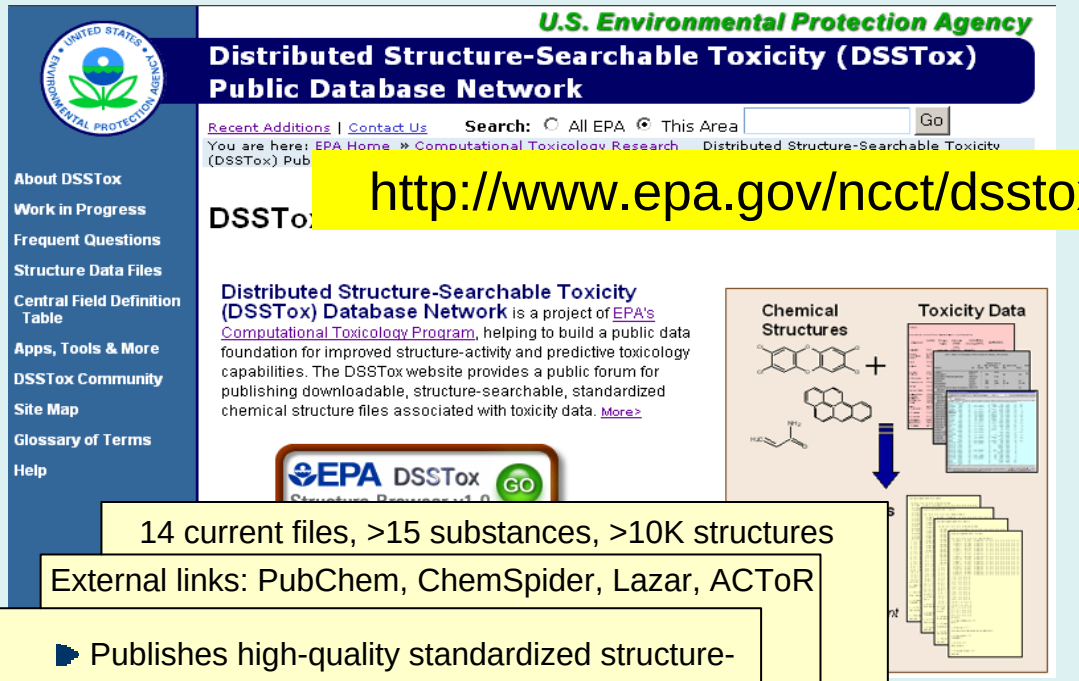
zed
data for use
odeling

DSSTox Project

&

Chemical QC

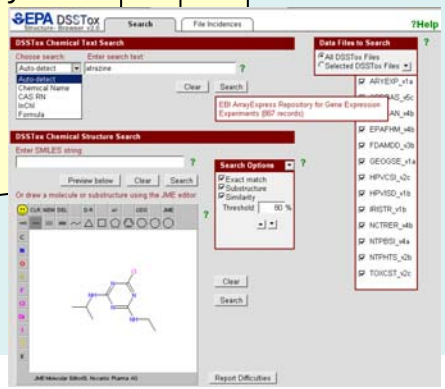
DSSTox: Distributed Structure-Searchable Database Network Project…& ToxCast™



14 current files, >15 substances, >10K structures

External links: PubChem, ChemSpider, Lazar, ACToR

- Publishes high-quality standardized structure-data (SD) files pertaining to toxicology:
  - *EPA, HPV-IS, IRIS, NTP, FDA, NCBI, EBI, …*
  - Strict chemical QA procedures
- Public substance/structure ID registry system
  - DSSTox Structure-Browser
    - *Search by structure, name, CAS,*
    - *Similarity & fragment search*
    - *Search within a DSSTox file*
    - *Link-in and link out of browser*



- Chemical information QA and structure annotation
  - *ToxCast Phase I & II*
  - *ToxRefDB*
  - *Tox21*
- Primary source of QC'd structures for ACToR
- Facilitate external linkages and data publication
- Publish summary activities and chemical classifiers for modeling
- PubChem Source depositor for structures & "assays"

DSSTox Standard Chemical Fields:

STRUCTURE

DSSTox_CID

STRUCTURE_Formula

STRUCTURE_MolecularWeight

STRUCTURE_ChemicalType

STRUCTURE_TestedForm_DefinedOrganic

STRUCTURE_ChemicalName_IUPAC

STRUCTURE_SMILES

STRUCTURE_Parent_SMILES

STRUCTURE_InChI

STRUCTURE_Shown

- tested chemical,
- general form of chemical,
- active ingredient of formulation,
- representative isomer in mixture,
- representative component in mixture,
- monomer of polymer

, simplified to parent

- defined organic
- inorganic
- organometallic

- parent,
- salt Na, Cl, etc
- complex HCl, H2O, mesylate, etc

DSSTox_Generic_SID

TestSubstance_ChemicalName

TestSubstance_CASRN

TestSubstance_Description

- single chemical compound
- macromolecule
- mixture or formulation
- unspecified or multiple forms

ChemicalNote

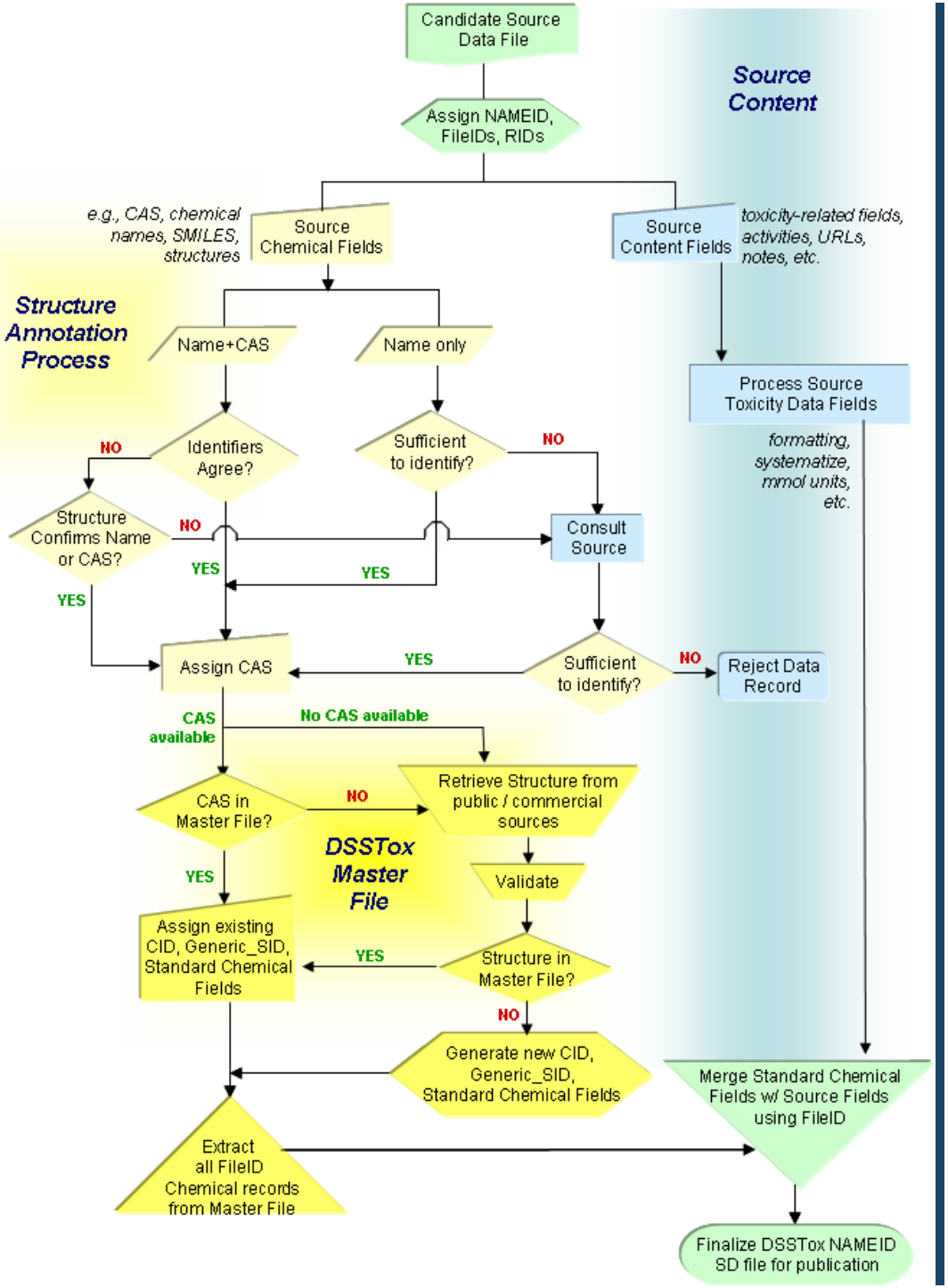
- description of mixture
- nature & CAS of components
- tautomers
- stereochemistry

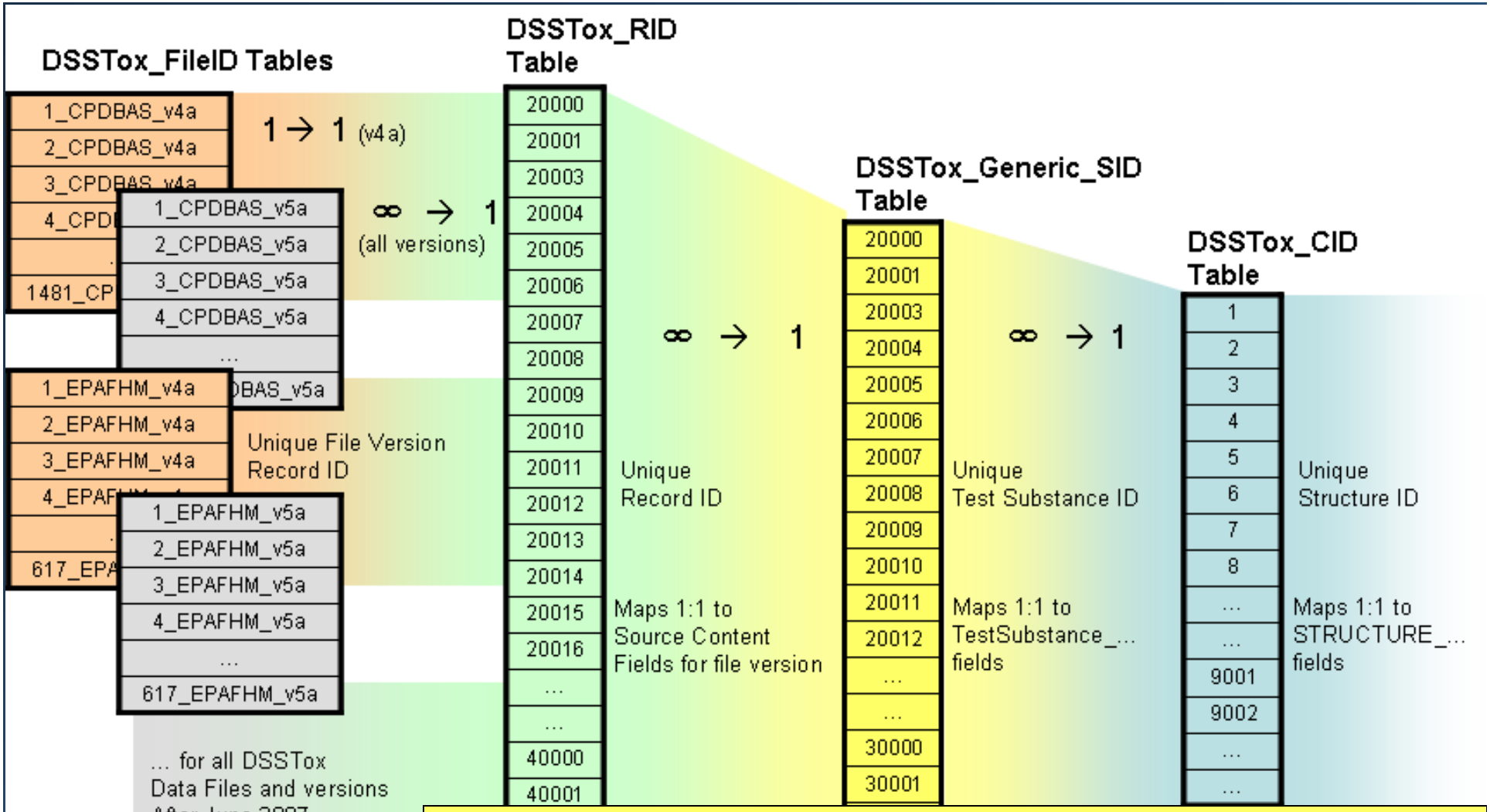
DSSTox_RID

DSSTox_FileID

DSSTox Chemical Quality Control Procedures:

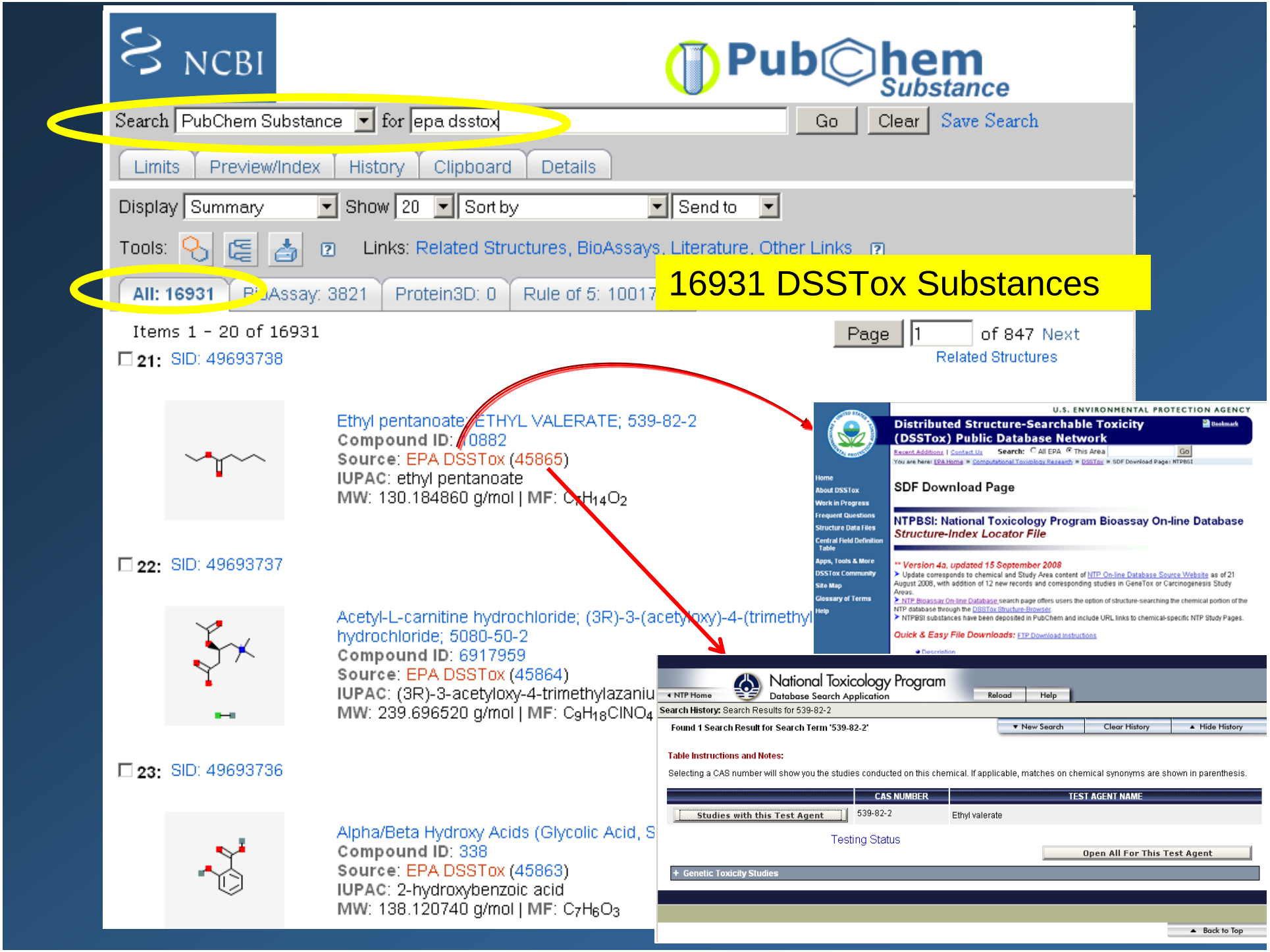
- Chemical identification
- Structure annotation
- Substance details
- Label consistency
- Internal consistency
- PubChem deposits

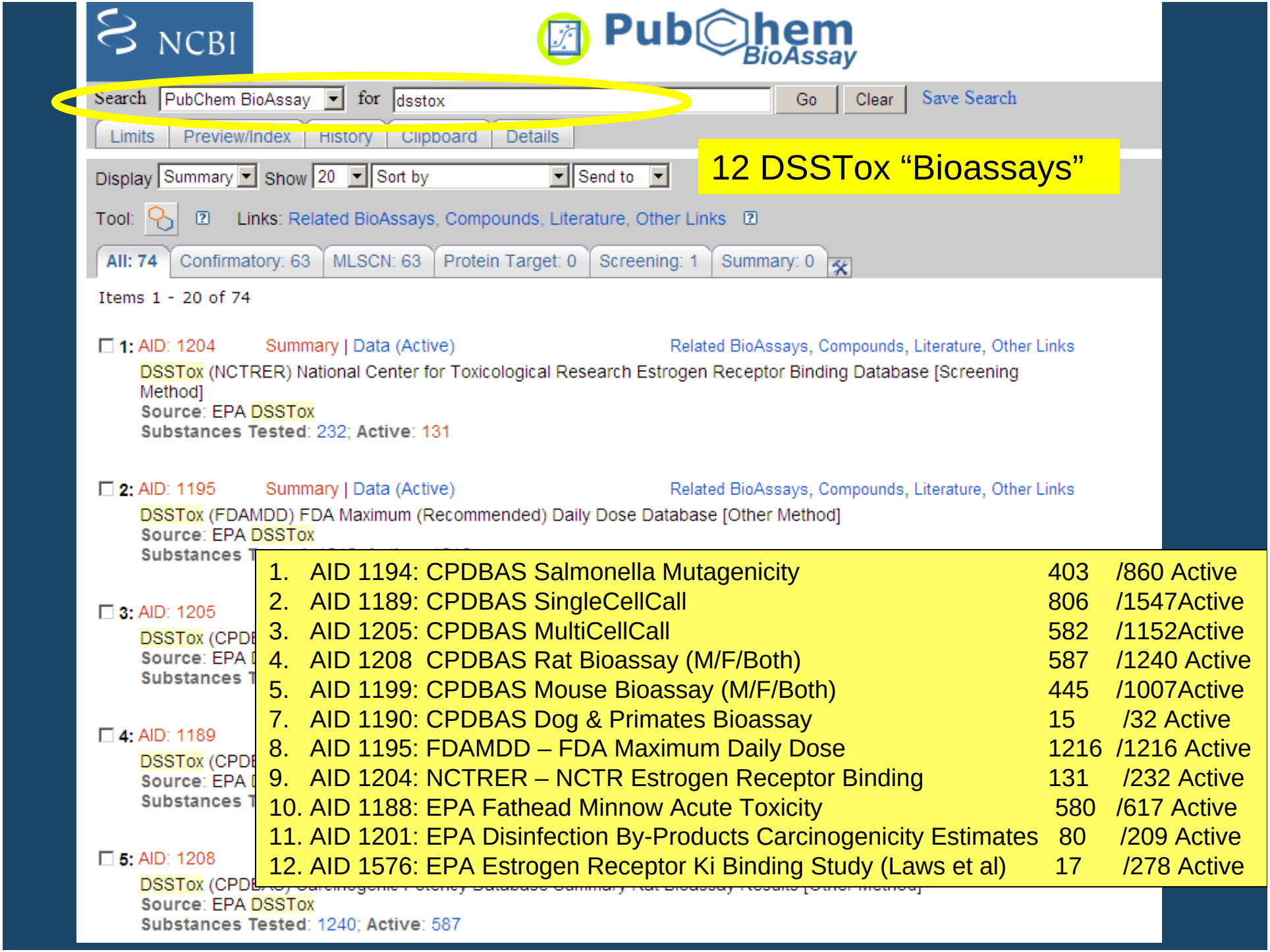




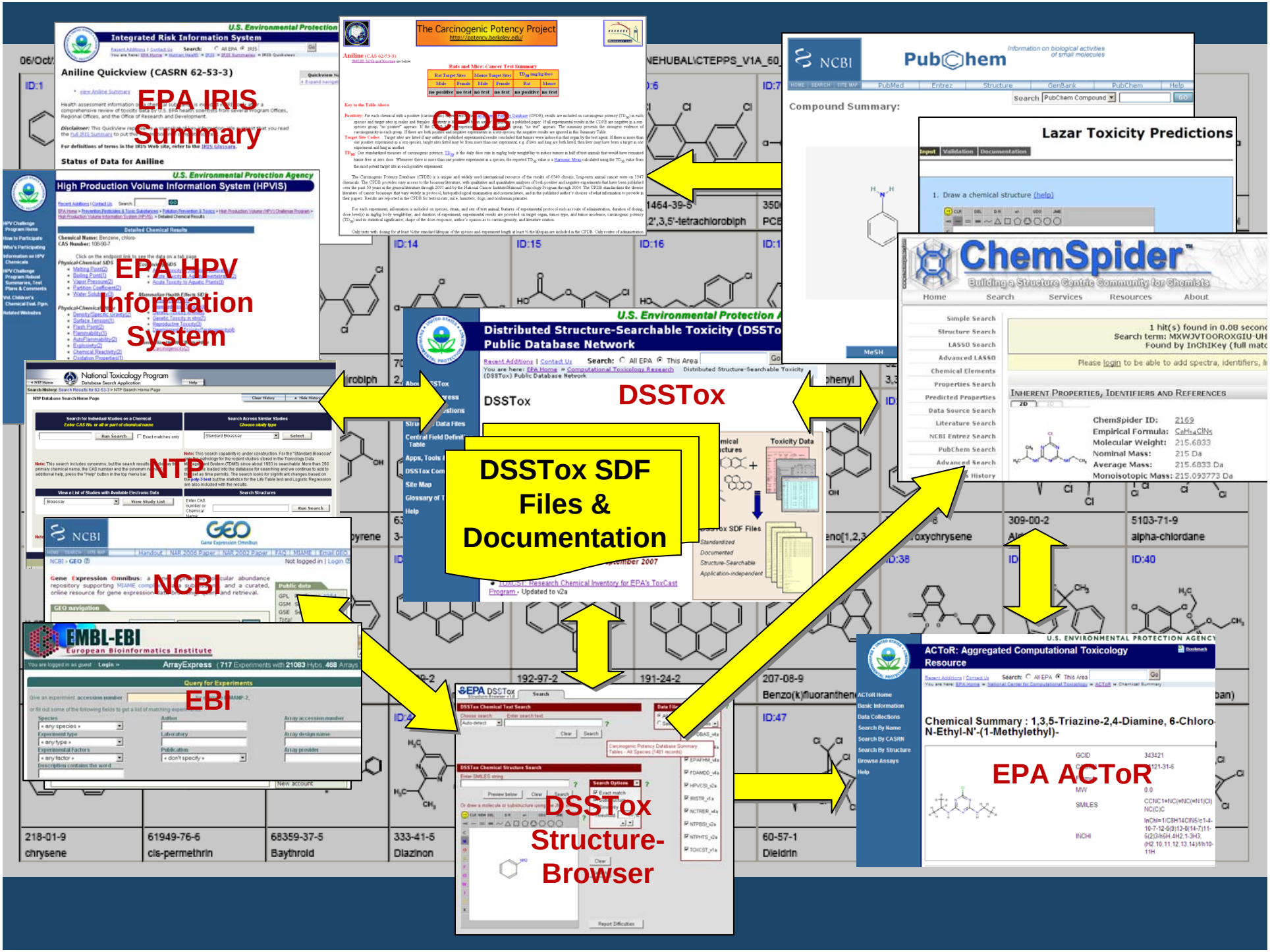
DSSTox Master File

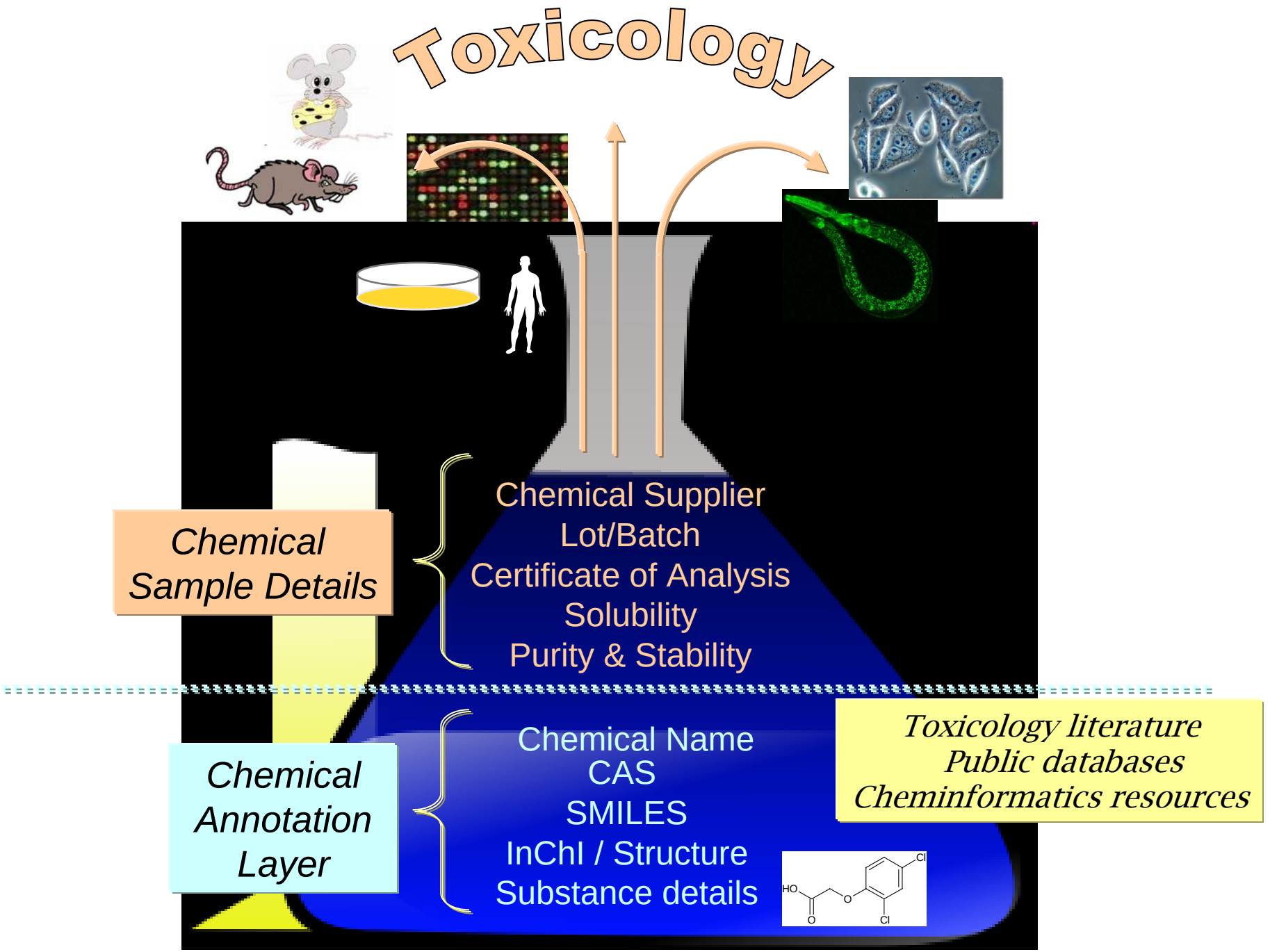
- Standardized representation of structures (CID) and substances (Generic_SID) across all files
- Public chemical registry system allows user to create database from published DSSTox files


NCBI
PubChem Substance
Search PubChem Substance for epa dsstox
Go
Clear
Save Search
Limits
Preview/Index
History
Clipboard
Details
Display Summary
Show 20
Sort by
Send to
Tools:
Links: Related Structures, BioAssays, Literature, Other Links
All: 16931
Protein3D: 0
Rule of 5: 10017
16931 DSSTox Substances
Items 1 - 20 of 16931
Page 1 of 847 Next
Related Structures
21: SID: 49693738
Ethyl pentanoate; ETHYL VALERATE; 539-82-2
Compound ID: 0882
Source: EPA DSSTox (45865)
IUPAC: ethyl pentanoate
MW: 130.184860 g/mol | MF: C7H14O2
22: SID: 49693737
Acetyl-L-carnitine hydrochloride; (3R)-3-(acetyloxy)-4-(trimethyl
hydrochloride; 5080-50-2
Compound ID: 6917959
Source: EPA DSSTox (45864)
IUPAC: (3R)-3-acetyloxy-4-trimethylazaniu
MW: 239.696520 g/mol | MF: C9H18ClNO4
23: SID: 49693736
Alpha/Beta Hydroxy Acids (Glycolic Acid, S
Compound ID: 338
Source: EPA DSSTox (45863)
IUPAC: 2-hydroxybenzoic acid
MW: 138.120740 g/mol | MF: C7H6O3
U.S. ENVIRONMENTAL PROTECTION AGENCY
Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network
SDF Download Page
NTPBSI: National Toxicology Program Bioassay On-line Database
Structure-Index Locator File
** Version 4a, updated 15 September 2008
Quick & Easy File Downloads:
National Toxicology Program
Database Search Application
Search History: Search Results for 539-82-2
Found 1 Search Result for Search Term '539-82-2'
New Search
Clear History
Hide History
Table Instructions and Notes:
Selecting a CAS number will show you the studies conducted on this chemical. If applicable, matches on chemical synonyms are shown in parenthesis.
CAS NUMBER
TEST AGENT NAME
Studies with this Test Agent
539-82-2
Ethyl valerate
Testing Status
Open All For This Test Agent
Genetic Toxicity Studies
Back to Top

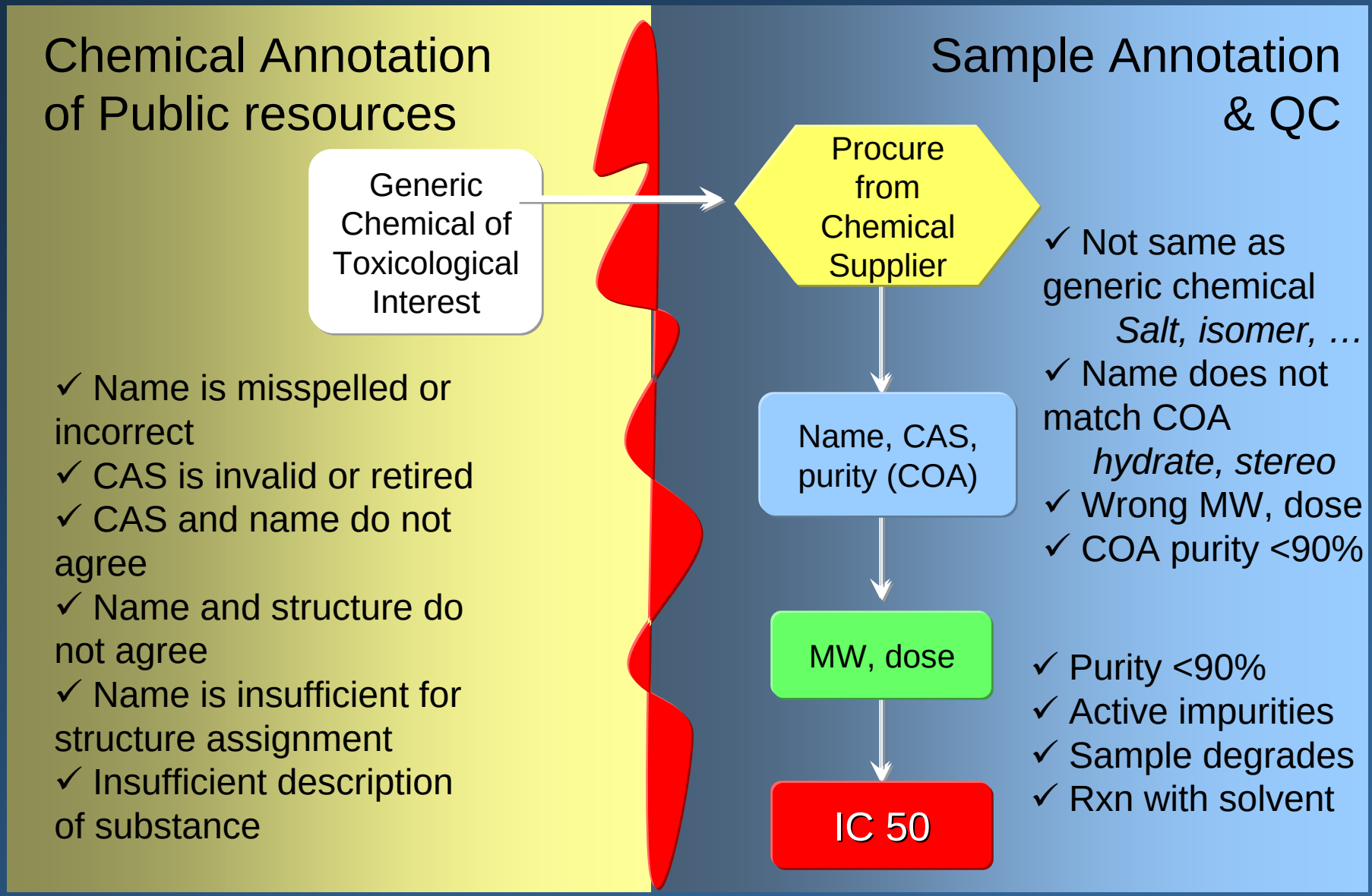


1.	AID 1194: CPDBAS Salmonella Mutagenicity	403	/860 Active
2.	AID 1189: CPDBAS SingleCellCall	806	/1547Active
3.	AID 1205: CPDBAS MultiCellCall	582	/1152Active
4.	AID 1208 CPDBAS Rat Bioassay (M/F/Both)	587	/1240 Active
5.	AID 1199: CPDBAS Mouse Bioassay (M/F/Both)	445	/1007Active
7.	AID 1190: CPDBAS Dog & Primates Bioassay	15	/32 Active
8.	AID 1195: FDAMDD – FDA Maximum Daily Dose	1216	/1216 Active
9.	AID 1204: NCTRER – NCTR Estrogen Receptor Binding	131	/232 Active
10.	AID 1188: EPA Fathead Minnow Acute Toxicity	580	/617 Active
11.	AID 1201: EPA Disinfection By-Products Carcinogenicity Estimates	80	/209 Active
12.	AID 1576: EPA Estrogen Receptor Ki Binding Study (Laws et al)	17	/278 Active


EPA IRIS Summary
CPDB
EPA HPV Information System
DSSTox
DSSTox SDF Files & Documentation
NTP
NCBI
EBI
DSSTox Structure-Browser
EPA ACToR
Integrated Risk Information System
Aniline Quickview (CASRN 62-53-3)
The Carcinogenic Potency Project
High Production Volume Information System (HPVIS)
Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network
National Toxicology Program
Gene Expression Omnibus
EMBL-EBI
European Bioinformatics Institute
ACToR: Aggregated Computational Toxicology Resource
PubChem
Lazar Toxicity Predictions
ChemSpider
Chemical Summary : 1,3,5-Triazine-2,4-Diamine, 6-Chloro-N-Ethyl-N'-(1-Methylethyl)-

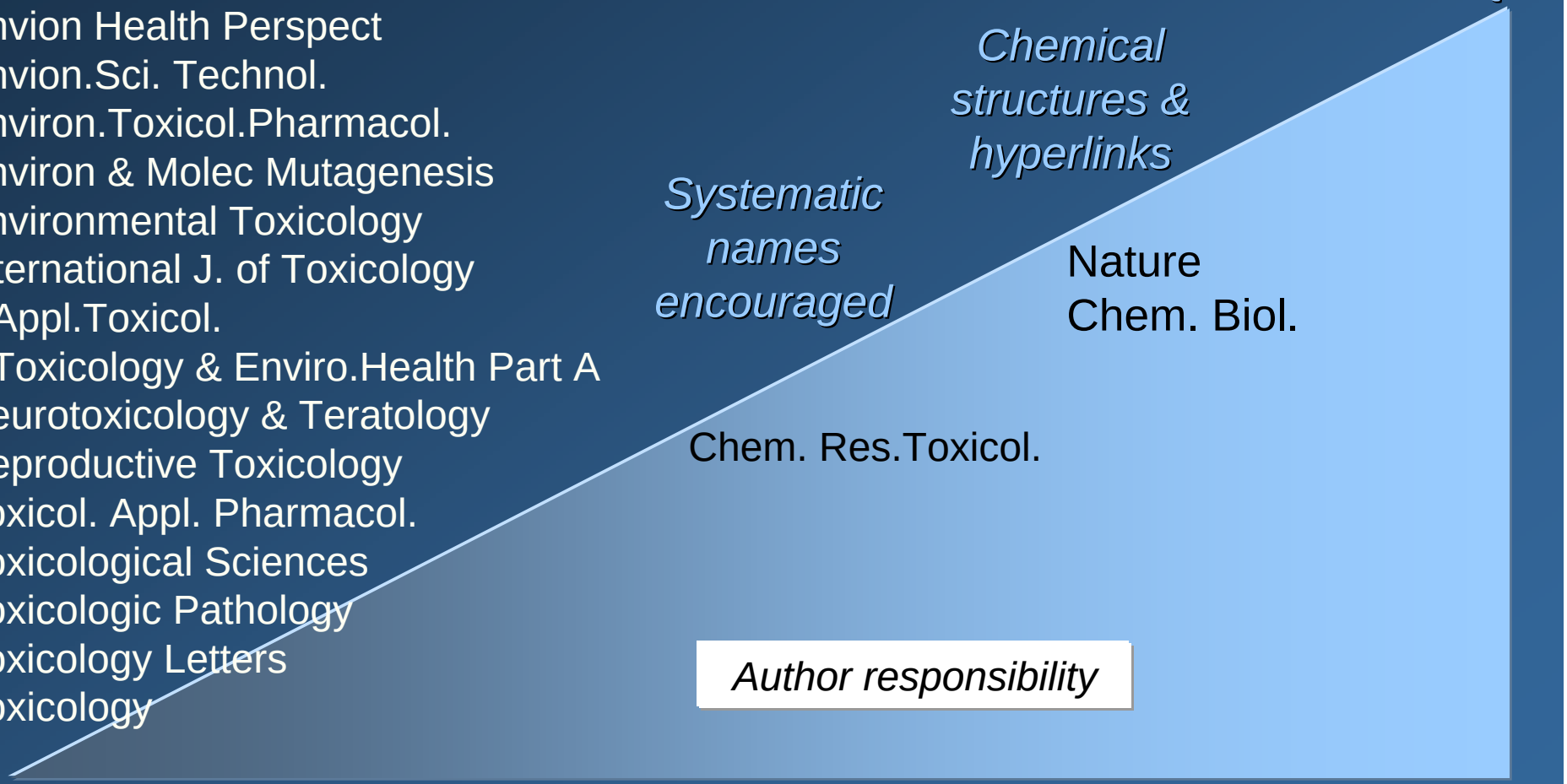

Toxicology
Chemical Sample Details
Chemical Supplier
Lot/Batch
Certificate of Analysis
Solubility
Purity & Stability
Chemical Annotation Layer
Chemical Name
CAS
SMILES
InChI / Structure
Substance details
Toxicology literature
Public databases
Cheminformatics resources
HO
O
O
Cl
Cl

Chemical Errors in Toxicity Information:



Current Guidelines & Chemical Standards in Toxicology & Biological Effects Publishing:

Archives of Toxicology
Brain Research Bulletin
Ecotoxicol. Environ. Saf.
Envion Health Perspect
Envion.Sci. Technol.
Environ.Toxicol.Pharmacol.
Environ & Molec Mutagenesis
Environmental Toxicology
International J. of Toxicology
J.Appl.Toxicol.
J Toxicology & Enviro.Health Part A
Neurotoxicology & Teratology
Reproductive Toxicology
Toxicol. Appl. Pharmacol.
Toxicological Sciences
Toxicologic Pathology
Toxicology Letters
Toxicology



"Authors should provide sufficient detail to allow the work to be reproduced."

Current Guidelines & Chemical Standards in Toxicology & Biological Effects Publishing:

Archives of Toxicology
Brain Research Bulletin
Ecotoxicol. Environ. Saf.
Envion Health Perspect
Envion.Sci. Technol.
Environ.Toxicol.Pharmacol.
Environ & Molec Mutagenesis
Environmental Toxicology
International J. of Toxicology
J.Appl.Toxicol.
J Toxicology & Enviro.Health
Neurotoxicology & Teratology
Reproductive Toxicology
Toxicol. Appl. Pharmacol.
Toxicological Sciences
Toxicologic Pathology
Toxicology Letters
Toxicology

In[illegible]endent Ch[illegible]cal QC

Survey journal articles over 4 year period:

- Approx 30 cases of reported purity (from Supplier)
- Aside from reports of new compounds (rare), virtually no analytical QC follow-up.

Author responsibility

"Authors should provide sufficient detail to allow the work to be reproduced."

nature chemical biology

Editorial

Nature Chemical Biology **3**, 297 (2007)
doi:10.1038/nchembio0607-297

A new look for chemical information

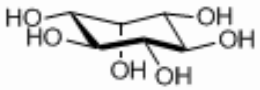
***Nature Chemical Biology* is committed to enhancing interdisciplinary communication and features online content to increase the accessibility of chemical information for our readers.**

"**Chemical compound information**
For all significant compounds included in Nature Chemical Biology original research papers, a compound data page, linked directly from the compound reference in the full text, appears in the online journal.

Compound Data Index

From the following article
Small molecules enhance autophagy and reduce toxicity in Huntington's disease models
Sovan Sarkar, Ethan O Perlstein, Sara Imarisio, Sandra Pineau, Axelle Cordenier, Rebecca L Maglathlin, John A Webster, Timothy A Lewis, Cahir J O'Kane, Stuart L Schreiber & David C Rubinsztein

Nature Chemical Biology **3**, 331-338 (2007) Published online: 7 May 2007
doi:10.1038/nchembio883



Compound 1
Inositol
View in PubChem
View compound page (2 KB) | View in 3D (2 KB) | Download ChemDraw file of structure (2 KB)

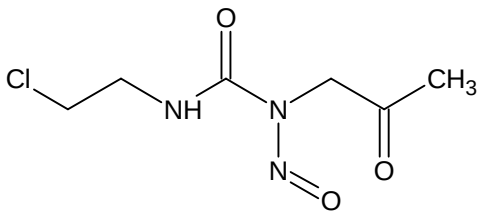
A PubChem link allows users of Nature Chemical Biology to go, in a single click, from the mention of a molecule in a paper to a rich and growing collection of information about chemical structures and their biological assay results, hosted by the NCBI.

"Authors ensure that the chemical compound information within their papers is complete, scientifically accurate and appropriately formatted."

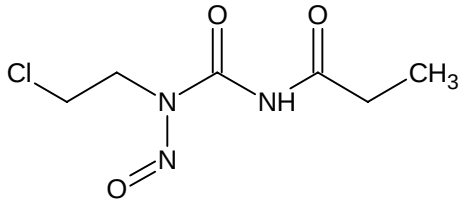
Carcinogenic Potency Database:
Hamster Carcinogenicity Results

From CPDB Hamster Table:

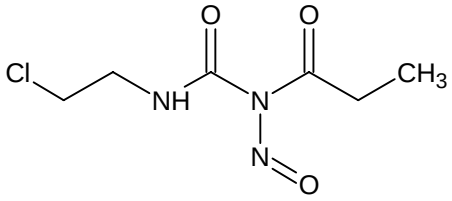
N-Nitroso-oxopropylchloroethylurea NOCAS



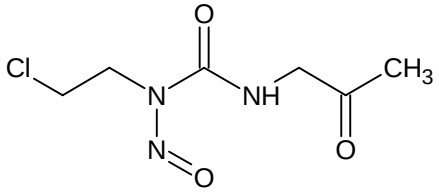
3-(2-chloroethyl)-1-nitroso-1-(2-oxopropyl)urea



N-{[(2-chloroethyl)(nitroso)amino]carbonyl}propanamide



N-{[(2-chloroethyl)amino]carbonyl}-*N*-nitrosopropanamide



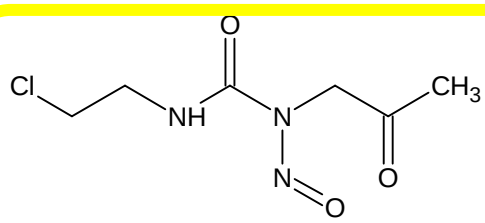
1-(2-chloroethyl)-1-nitroso-3-(2-oxopropyl)urea

CPDB Hamster Carcinogenicity Data Reference → Rat Cancer Study (Materials) → Chemical synthesis paper, NMR, IR structure confirmation

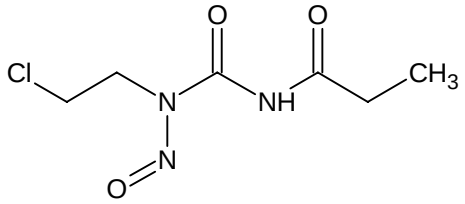
Carcinogenic Potency Database:
Hamster Carcinogenicity Results

From CPDB Hamster Table:

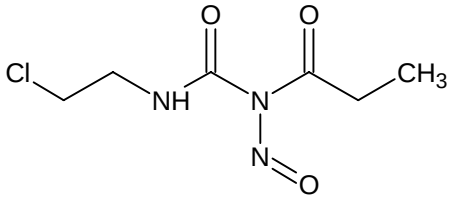
N-Nitroso-oxopropylchloroethylurea NOCAS



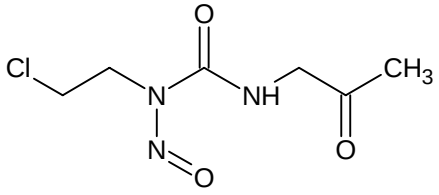
3-(2-chloroethyl)-1-nitroso-1-(2-oxopropyl)urea



N-{[(2-chloroethyl)(nitroso)amino]carbonyl}propanamide



N-{[(2-chloroethyl)amino]carbonyl}-*N*-nitrosopropanamide



1-(2-chloroethyl)-1-nitroso-3-(2-oxopropyl)urea

Chemical synthesis paper, NMR, IR structure confirmation → IUPAC Name → CAS →

Determined to be same chemical as:
1-(2-Oxopropyl)nitroso-3-(2-chloroethyl)urea
CAS [110559-85-8]
Already listed in CPDB Rat and Mouse Table

Perfluoroalkylacids (PFAAs):

- Man-made, lipophilic, stable, biopersistant
- Widespread industrial use as surfactants *(stain and oil-resistant coatings, microwave popcorn bags, emulsifier, etc)*
- Widespread exposure & environmental contamination
- PFOA (perflurooctanoic acid ammonium salt) and PFOS (perfluorooctane sulfonic acid) of greatest health concern
- PFOA and PFOS have undergone extensive toxicity testing
  - ***Hepatotoxic***
  - ***Developmental toxicants***
  - ***Immunotoxic***

About • Contact • Search

National Library of Medicine
Specialized Information Services

SIS NLM

ChemIDplus Lite
Full Record

Tox. & Env. Health | TOXNET | Return to Results Page

Potassium perfluorooctanesulfonate
RN: 2795-39-3

Names and Synonyms

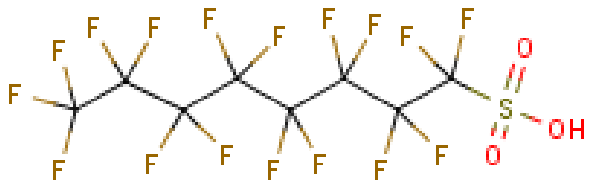
Synonyms
- 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-Heptadecafluoro-1-octanesulfonic acid, potassium salt
- 1-Octanesulfonic acid, heptadecafluoro-, potassium salt
- AI3-50950
- EINECS 220-527-1
- FC 95
- Floral FC 95
- Fluorad FC 95
- Heptadecafluorooctanesulfonic acid, potassium salt
- NSC 18405
- Perfluorooctanesulfonic acid, potassium salt
- Potassium PFOS
- Potassium heptadecafluorooctane-1-sulfonate
- Potassium perfluorooctanesulfonate

Systematic Name
- 1-Octanesulfonic acid, 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluoro-, potassium salt
- Potassium heptadecafluorooctane-1-sulphonate

Registry Numbers

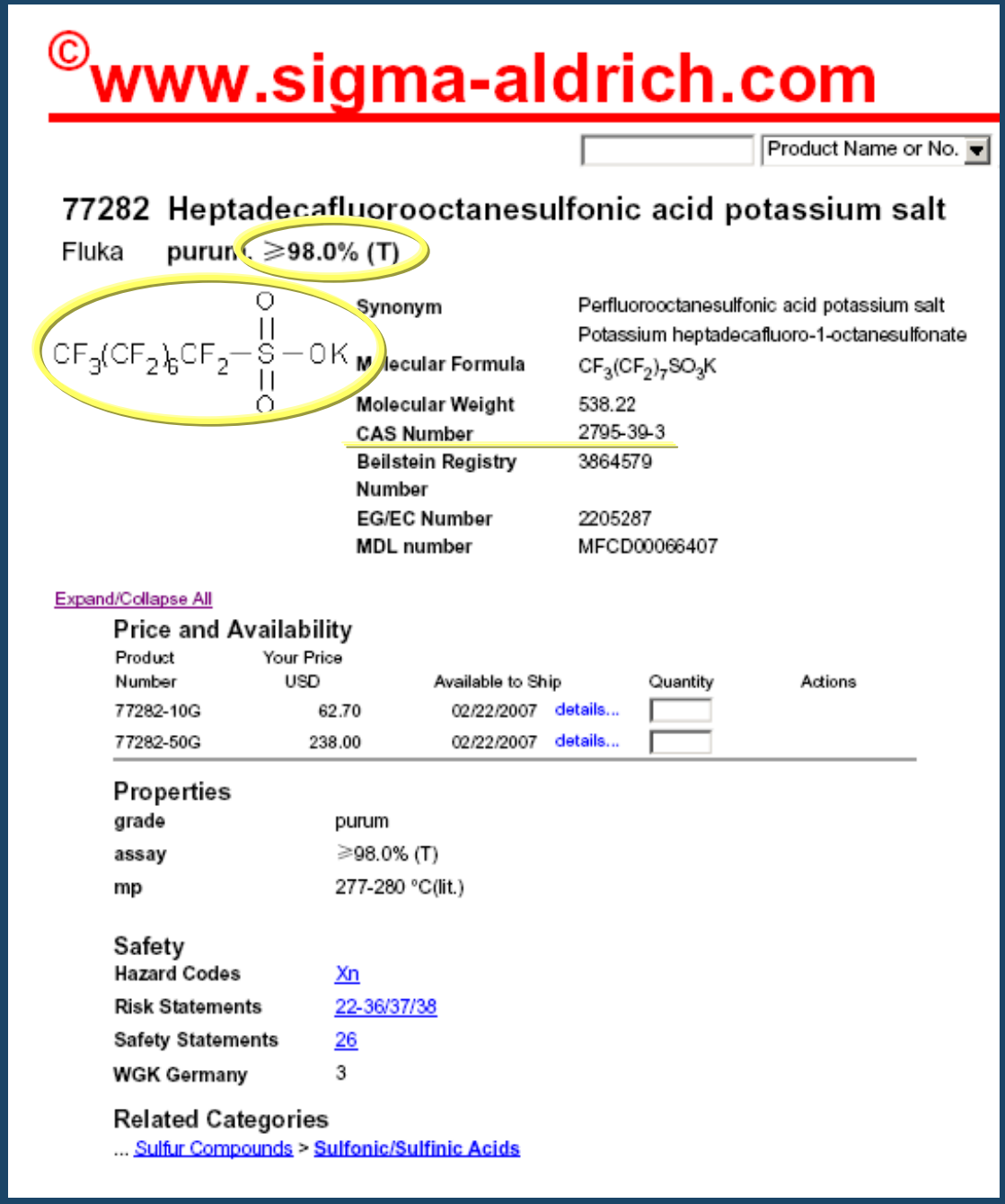
CAS Registry Number
- 2795-39-3

Other Registry Number
- 117925-64-1
- 59112-13-9
- 62010-27-9
- 69458-54-4



PFOS

- Major synthetic pathway is electrochemical fluorination
- Yields approx 60-80% linear form, with significant non-linear (branched form) contamination
- Verified by primary manufacturer (3M) and NMR



PFOS

- Listed as 98.0% pure (T)
- (T) indicates titration method, which confirms only empirical formula
- Listed incorrectly as linear form by both structure and CASRN
- No Certificate of Analysis available from Sigma (Fluka)

Problems with using CAS as Primary Registry for Public Toxicity Databases (and EPA):

CAS Registry Number

3068-88-0 REGISTRY

Deleted Registry Number

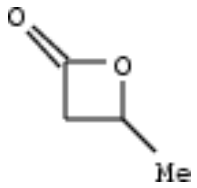
36536-46-6, 43137-57-1

Chemical Name

2-Oxetanone, 4-methyl- (CA INDEX NAME)
Butyric acid, .beta.-hydroxy-, lactone (4CI)
Butyric acid, 3-hydroxy-, .beta.-lactone (6CI)
(.+-.)-.beta.-Butyrolactone
(.+-.)-.beta.-Methylpropiolactone
(RS)-.beta.-Butyrolactone
.beta.-Butyrolactone
.beta.-Methyl-.beta.-propiolactone
.beta.-Methylpropiolactone
3-Hydroxybutyric acid lactone
4-Methyl-2-oxetanone
Butanoic acid, 3-hydroxy-, .beta.-lactone
DL-.beta.-Butyrolactone

Molecular Formula

C4 H6 O2



- Commercial, costly, unavailable to many
- High incidence of errors in assigning CAS with chemical and data in tox literature (1-10%)
- CAS assigned post-study by biologists/toxicologists
- Test substance annotation unavailable for purity grade, mixture details, etc.
- Not unique in public resources – deleted, alternate

Chemical Annotation of Public MicroArray Resources:

GEO & ArrayExpress

Towards a Public “Toxico-chemogenomics” Capability

Chemical Names:

- Non-standard entry
- No ability to structure-search or structure-analog search
- Fewer than 30 CAS in >2000 experiments

Accession Number	Listed Chemical Name
E-MEXP-132	estrogen
E-MAXD-39	oestradiol
E-GEOD-4025	estradiol (E2)
E-GEOD-848	E2
E-NASC-65	estradiol
E-SMDB-1443	Beta-estradiol
E-TABM-269	Estradiol
E-AFMX-12	17 beta-estradiol (E2)
E-AFMX-13	Estrogen
E-GEOD-1045	Estradiol
E-GEOD-1153	Estradiol E2
E-GEOD-2195	17beta-estradiol
E-GEOD-2251	17beta-estradiol
E-GEOD-2292	17beta-estradiol
E-GEOD-2889	17beta-estradiol
E-GEOD-3529	Estradiol(E2)
E-GEOD-4668	17beta-estradiol
E-MEXP-1053	Estradol
E-TABM-231	Estradiol

Abbreviations:

CCCP …..
Carbonyl Cyanide 3-ChloroPhenylhydrazone
Cat Colony Care Programme
Caltech Core Collapse Project
Mtm or MMC ….. Mitomycin C
DOX …... Doxorubicin or Doxycline
109 …… ?? triazine compound

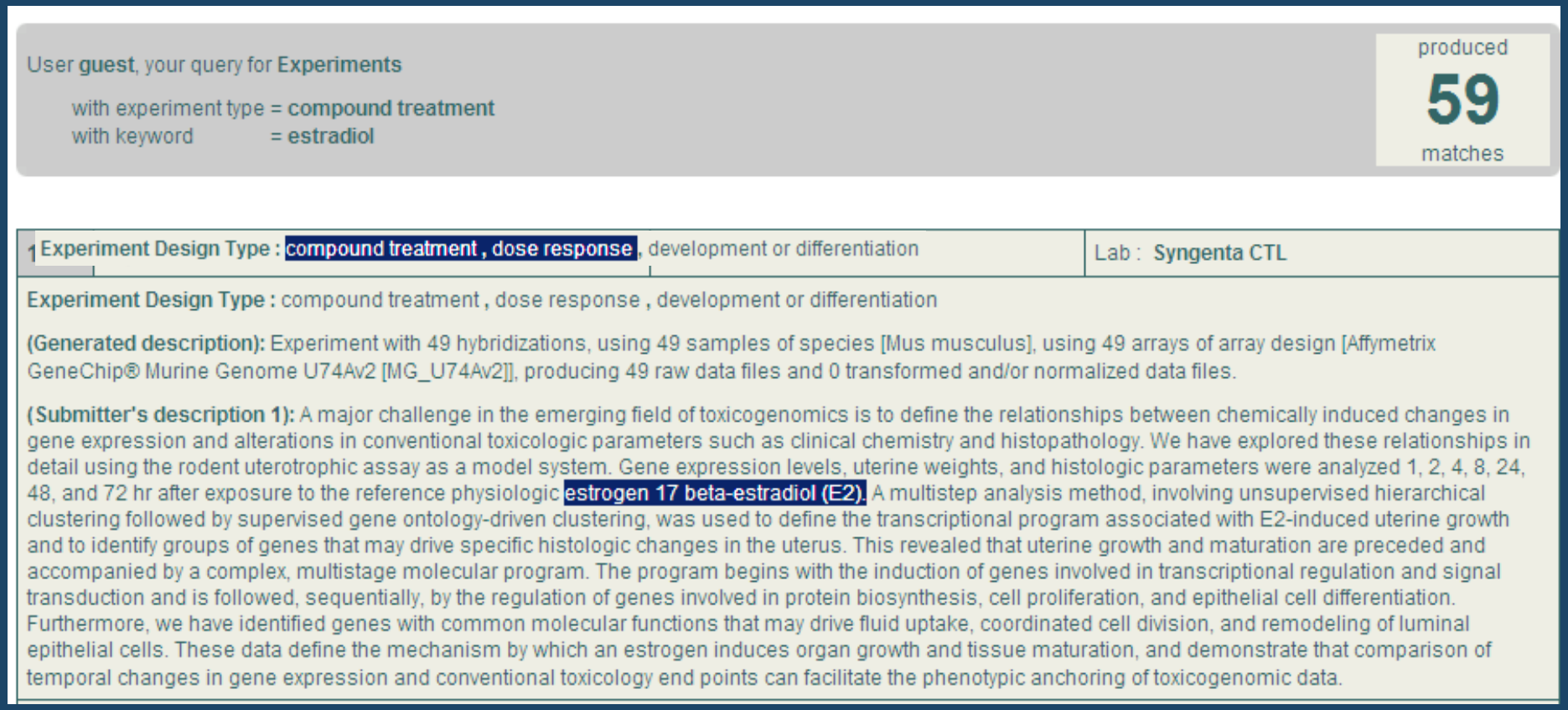
Incomplete Chemical names

Dipyridyl- A …… 4,4'-; 3,3'-; 2,4'-; 2,3'-; or 2,2'- ?
Succinate …… Na, K, Ca?

Misspellings:

Carconyl chlotide … Carbonyl chloride (phosgene)
ciprofbrate …. Ciprofibrate

EBI ArrayExpress: Submitter Experiment Description



User **guest**, your query for **Experiments**

with experiment type = **compound treatment**
with keyword = **estradiol**

produced **59** matches

1 **Experiment Design Type :** compound treatment , dose response , development or differentiation — Lab : **Syngenta CTL**

Experiment Design Type : compound treatment , dose response , development or differentiation

(Generated description): Experiment with 49 hybridizations, using 49 samples of species [Mus musculus], using 49 arrays of array design [Affymetrix GeneChip® Murine Genome U74Av2 [MG_U74Av2]], producing 49 raw data files and 0 transformed and/or normalized data files.

(Submitter's description 1): A major challenge in the emerging field of toxicogenomics is to define the relationships between chemically induced changes in gene expression and alterations in conventional toxicologic parameters such as clinical chemistry and histopathology. We have explored these relationships in detail using the rodent uterotrophic assay as a model system. Gene expression levels, uterine weights, and histologic parameters were analyzed 1, 2, 4, 8, 24, 48, and 72 hr after exposure to the reference physiologic estrogen 17 beta-estradiol (E2). A multistep analysis method, involving unsupervised hierarchical clustering followed by supervised gene ontology-driven clustering, was used to define the transcriptional program associated with E2-induced uterine growth and to identify groups of genes that may drive specific histologic changes in the uterus. This revealed that uterine growth and maturation are preceded and accompanied by a complex, multistage molecular program. The program begins with the induction of genes involved in transcriptional regulation and signal transduction and is followed, sequentially, by the regulation of genes involved in protein biosynthesis, cell proliferation, and epithelial cell differentiation. Furthermore, we have identified genes with common molecular functions that may drive fluid uptake, coordinated cell division, and remodeling of luminal epithelial cells. These data define the mechanism by which an estrogen induces organ growth and tissue maturation, and demonstrate that comparison of temporal changes in gene expression and conventional toxicology end points can facilitate the phenotypic anchoring of toxicogenomic data.

- Free text description
- No curation or external review
- No chemical indexing of experiment

Series GSE2187 Query DataSets for GSE2187

Status	Public on May 01, 2005
Title	Classification of a large micro-array dataset. Algorithm comparison and analysis of drug signatures.
Organism(s)	Rattus norvegicus
Summary	Classification of a large micro-array d… analysis of drug signatures. These data support the publication ti… array dataset. Algorithm comparison … Some of the calculations in the public… version of the data available at http:/… Copyright (c) 2005 by Iconix Pharmac… Guidelines for commercial use: http://www.iconixbiosciences.com/gu… Keywords: other
Contributor(s)	Natsoulis G, El Ghaoui L, Lanckriet GR, Tolley AM, Leroy F, Dunlea S, Eynon BP, Pearson CI, Tugendreich S, Jarnagin K
Citation(s)	Natsoulis G, El Ghaoui L, Lanckriet GR, Tolley AM et al. Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. *Genome Res* 2005 M…
Submission date	Jan 25, 2005
Contact name	Mark Fielden
Organization name	Iconix Biosciences
Street address	325 East Middlefield Road
City	Mountain View
State/province	CA
ZIP/Postal code	94043
Country	USA
Platforms (1)	GPL1820 Rat Uniset 10K
Samples (587)	GSM43278 1-NAPHTHYL ISOTHI…OIL_ORAL GAVAGE_RATM, Replicate1
± More…	GSM43279 1-NAPHTHYL ISOTHIOCYANATE_30_.25_LIVER_CORN OIL_ORAL GAVAGE_RATM, Replicate2

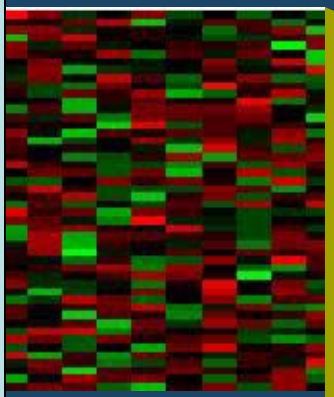
- No chemical standards
- Difficult to identify chemical exposure-related experiments
- Fewer than 30 CAS RNs provided in >2000 experiments

- PERL scripts to filter web-accessed data content
- Manual review of Submitter textual descriptions
- Creation of initial chemical & experimental index
- QC & structure annotation

NCBI

GEO Gene Expression Omnibus

9957
Experiments



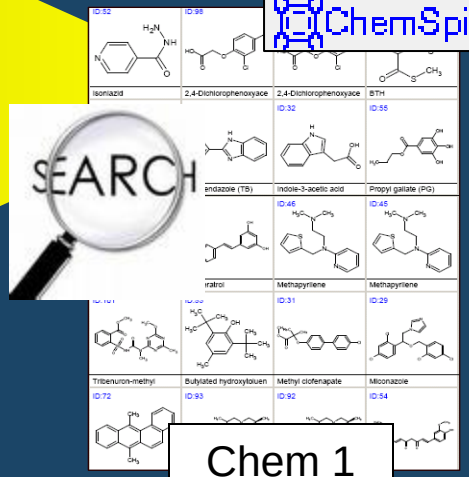
Experiments

Genes

Pathways

PubChem

ChemSpider



2381
Chemical experiments

Vehicle
Media
Reference
Combination

2134
Chemical treatment-related experiments

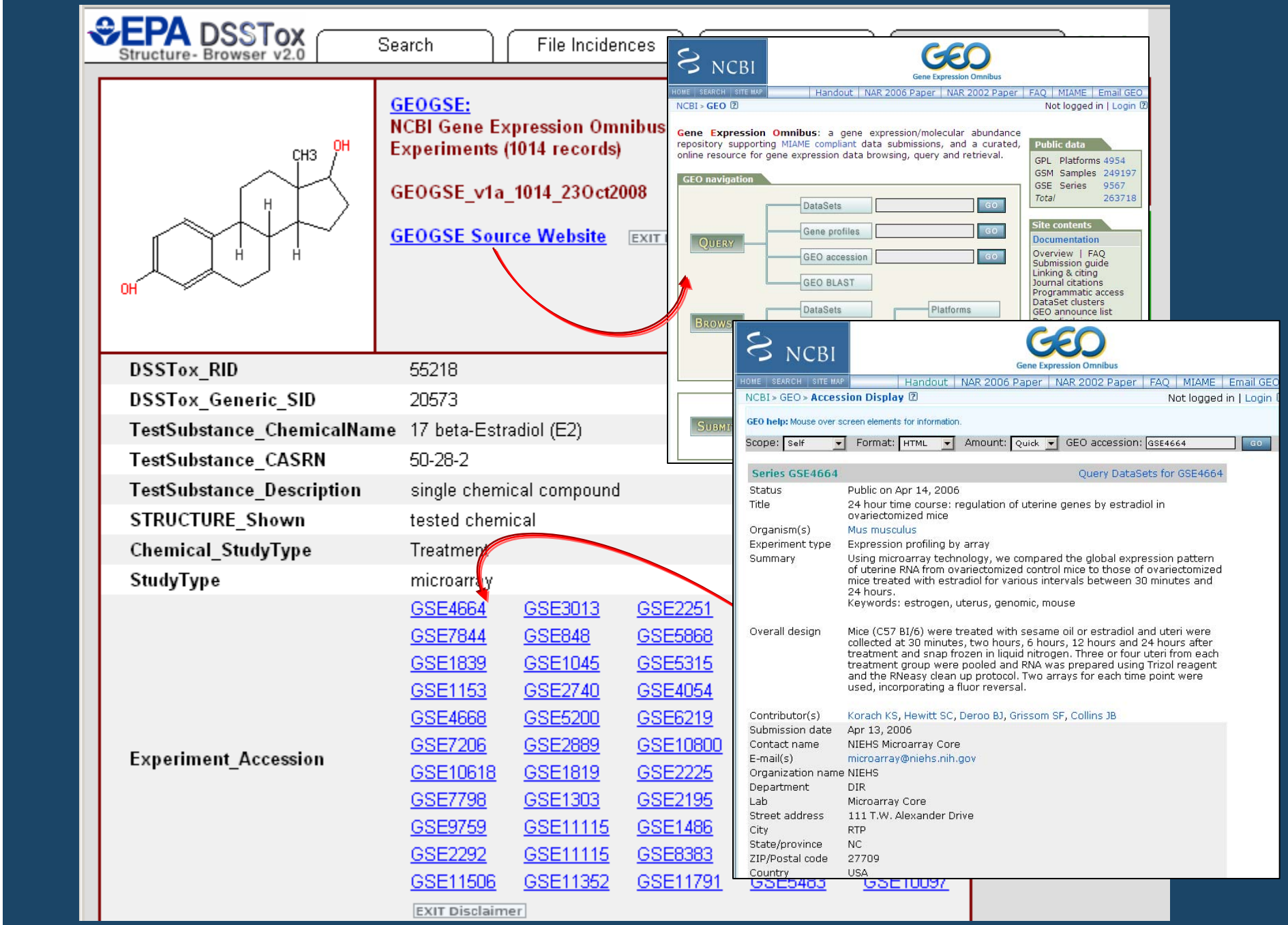
Treatment

1014
Unique chemicals

837 defined organics
84 inorganics
22 organometallics

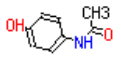
Chem 1
Chem 2
Chem 3
Chem 4
…
…
Chem 1014

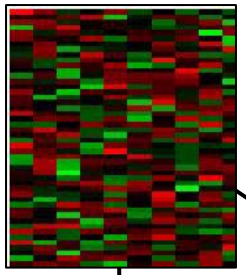
DSSTox
GEOCSI


EPA DSSTox
Structure- Browser v2.0
Search
File Incidences
CH3
OH
H
H
H
OH
GEOGSE:
NCBI Gene Expression Omnibus
Experiments (1014 records)
GEOGSE_v1a_1014_23Oct2008
GEOGSE Source Website
EXIT
DSSTox_RID 55218
DSSTox_Generic_SID 20573
TestSubstance_ChemicalName 17 beta-Estradiol (E2)
TestSubstance_CASRN 50-28-2
TestSubstance_Description single chemical compound
STRUCTURE_Shown tested chemical
Chemical_StudyType Treatment
StudyType microarray
Experiment_Accession
GSE4664 GSE3013 GSE2251
GSE7844 GSE848 GSE5868
GSE1839 GSE1045 GSE5315
GSE1153 GSE2740 GSE4054
GSE4668 GSE5200 GSE6219
GSE7206 GSE2889 GSE10800
GSE10618 GSE1819 GSE2225
GSE7798 GSE1303 GSE2195
GSE9759 GSE11115 GSE1486
GSE2292 GSE11115 GSE8383
GSE11506 GSE11352 GSE11791
GSE5463 GSE10097
EXIT Disclaimer
NCBI
Gene Expression Omnibus
HOME SEARCH SITE MAP
Handout NAR 2006 Paper NAR 2002 Paper FAQ MIAME Email GEO
NCBI > GEO
Not logged in | Login
Gene Expression Omnibus: a gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.
Public data
GPL Platforms 4954
GSM Samples 249197
GSE Series 9567
Total 263718
GEO navigation
QUERY
DataSets GO
Gene profiles GO
GEO accession GO
GEO BLAST
BROWSE
DataSets
Platforms
SUBMIT
Site contents
Documentation
Overview | FAQ
Submission guide
Linking & citing
Journal citations
Programmatic access
DataSet clusters
GEO announce list
NCBI
Gene Expression Omnibus
HOME SEARCH SITE MAP
Handout NAR 2006 Paper NAR 2002 Paper FAQ MIAME Email GEO
NCBI > GEO > Accession Display
Not logged in | Login
GEO help: Mouse over screen elements for information.
Scope: Self Format: HTML Amount: Quick GEO accession: GSE4664 GO
Series GSE4664
Query DataSets for GSE4664
Status Public on Apr 14, 2006
Title 24 hour time course: regulation of uterine genes by estradiol in ovariectomized mice
Organism(s) Mus musculus
Experiment type Expression profiling by array
Summary Using microarray technology, we compared the global expression pattern of uterine RNA from ovariectomized control mice to those of ovariectomized mice treated with estradiol for various intervals between 30 minutes and 24 hours.
Keywords: estrogen, uterus, genomic, mouse
Overall design Mice (C57 BI/6) were treated with sesame oil or estradiol and uteri were collected at 30 minutes, two hours, 6 hours, 12 hours and 24 hours after treatment and snap frozen in liquid nitrogen. Three or four uteri from each treatment group were pooled and RNA was prepared using Trizol reagent and the RNeasy clean up protocol. Two arrays for each time point were used, incorporating a fluor reversal.
Contributor(s) Korach KS, Hewitt SC, Deroo BJ, Grissom SF, Collins JB
Submission date Apr 13, 2006
Contact name NIEHS Microarray Core
E-mail(s) microarray@niehs.nih.gov
Organization name NIEHS
Department DIR
Lab Microarray Core
Street address 111 T.W. Alexander Drive
City RTP
State/province NC
ZIP/Postal code 27709
Country USA

Chemical Search Paradigm

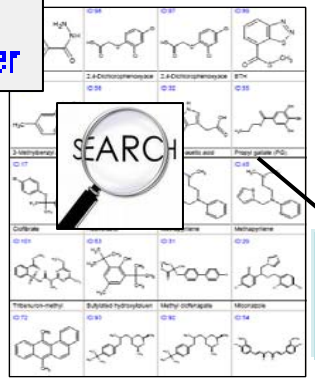
United States Environmental Protection Agency

? DSSTox Substance ID	Similarity Score%	Structure Match	Substance Name	CASRN	Substance Description	Details (Data Files)
20006	100		Acetaminophen	103-90-2	single chemical compound	CPDBAS EPAFHM FDAMDD NTPBSI NTPHTS



PubChem

ChemSpider



Toxicity Results

Bioassay Results

GEO Gene Expression Omnibus

ARRAYEXPRESS

GEO Results

200005594:Agilent;Rat
200005593:Agilent;Rat
200005595:Agilent;Rat
200000633:Custom Array; Rat
200005652:Agilent; Rat
200004874:Custom; Mouse
200005860:Agilent; Rat
200008858: GE Healthcare;Rat

ArrayExpress Results

E-TABM-131:Custom Array; Rat
E-MEXP-82: Affymetrix; Rat
E-TOXM-18: Agilent; Mouse
E-TOXM-31:Custom Array ; Human

Acetaminophen Meta-DataSet

Conclusions from GEO/ArrayExpress Effort:

- Past attention to chemical aspects of experiments and data – none !
- Structure annotation and linkages enhances scientific value of resources
- Automated text mining methods inadequate; required significant manual review & curation
- Recommended addition of two required data-submitter fields:
  - *Unambiguous chemical name, CAS if available*
  - *Purpose of chemical in relation to experiment, select one: treatment, vehicle, reference, other.*

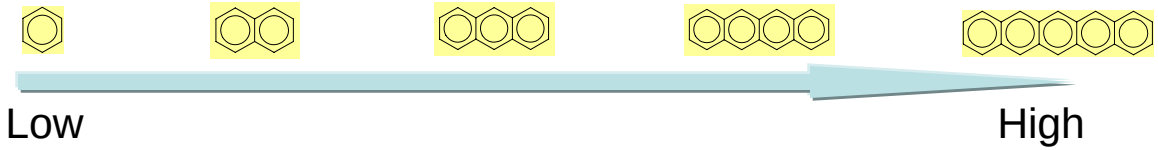
Recommendations:

- Toxicology journals & public databases should strengthen standards for chemical annotation and sample QC reporting
- ACS, CAS/STN should support public efforts to QC association of biological effects data with CAS/name/structure
  - *Allow >10K chemical substance/CAS lists for high-interest chemicals*
  - *Make STN available at no-low cost to public QC efforts*
- Chemical annotation efforts first step, but insufficient without chemical QC review
- Public database projects can contribute to QC
  - *Toxicology: DSSTox*
  - *ChemSpider – Wiki*
  - *PubChem – Source reliability for QC'd chemical info*

“Cheminformaticon”: Intrinsic Information Content

I. Representations of chemicals in public literature & databases

Intrinsic information content



Requires

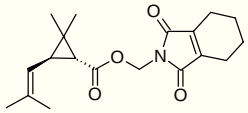
Examples

CASRN [7696-12-0]

Chemical Name
CCRIS 3284 Tetramethrin
(1,3-dioxo-1,3,4,5,6,7-hexahydro-2H-isoindol-2-yl)methyl (1R,3R)-2,2-dimethyl-3-(2-methylprop-1-en-1-yl)cyclopropanecarboxylate

SMILES
C\C(C)=C/[C@@H]3[C@@H](C(=O)OCN2C(=O)C=1CCCCC=1C2=O)C3(C)C

InChI
InChI=1/C19H25NO4/c1-11(2)9-14-15(19(14,3)4)18(23)24-10-20-16(21)12-7-5-6-8-13(12)17(20)22/h9,14-15H,5-8,10H2,1-4H3/t14-,15+/m1/s1



Chemical Structure

CAS → Name, Structure

Name → Structure

SMILES → Structure

InChI → Structure

Mol File

Acknowledgements:

- **EPA NCCT DSSTox Team:**
  Maritja Wolf (DSSTox) and Tom Transue (Structure-browser) – Lockheed Martin, Contractors to the US EPA
- **Toxicogenomics:** ClarLynda Williams (EPA)
- **EPA NCCT ToxCast Team:**
  Robert Kavlock (Director, NCCT)
  David Dix (ToxRefDB, HTS, Genomics)
  Keith Houck (HTS)
  Matt Martin (ToxRefDB)
  Richard Judson (ACToR, ToxMiner)
- **Literature/Journal review:**
  Inthirany Thillainadarajak (EPA SEEP)

This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.

“An expert is a person who has made all the mistakes that can be made in a very narrow field.”

– Niels Bohr

“Whoever is careless with the truth in small matters cannot be trusted with important matters.”

– Albert Einstein