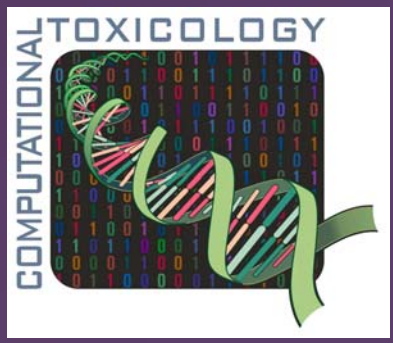
Vanderbilt Institute of Chemical Biology Seminar - March 28, 2007:

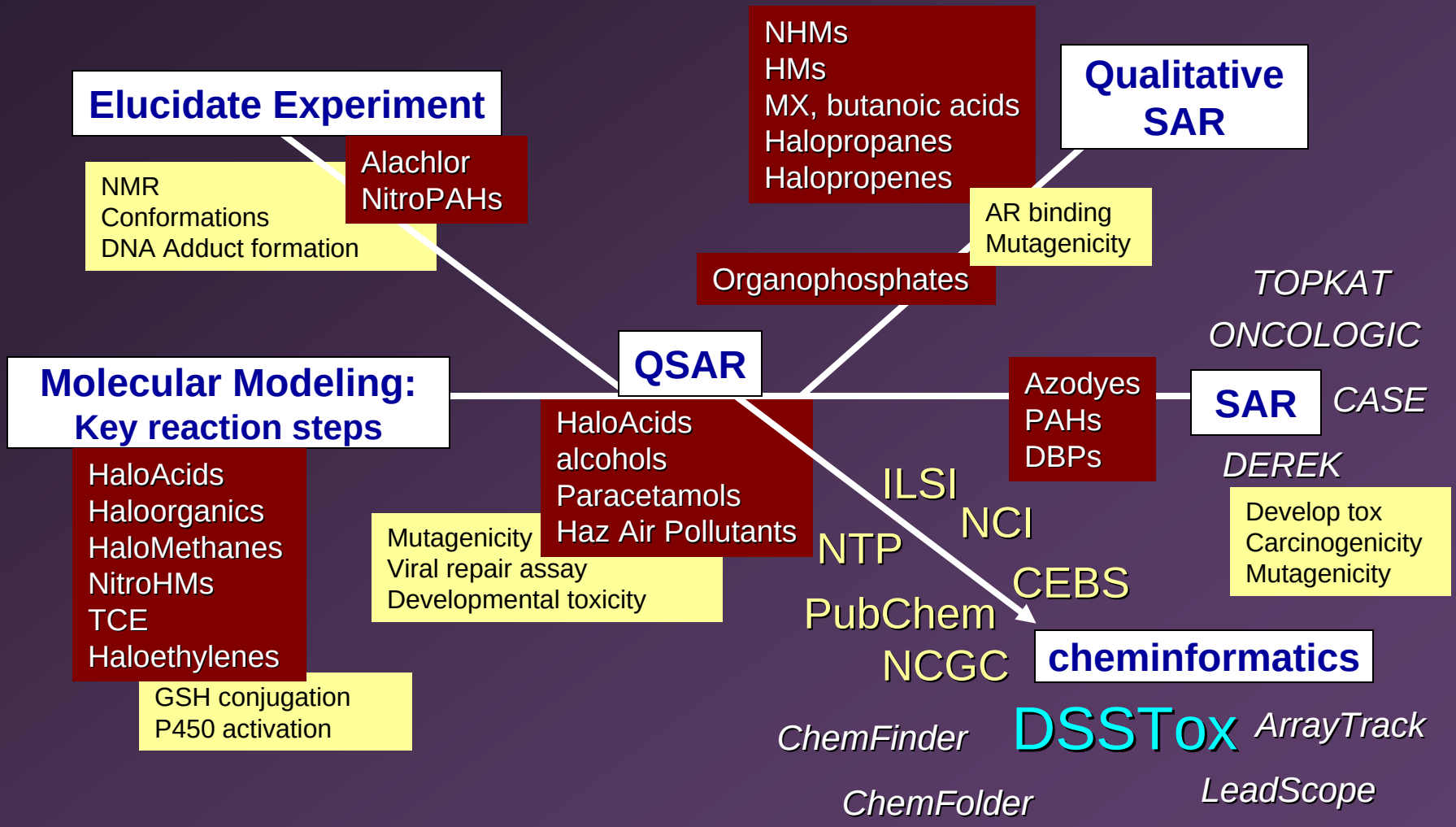
Expanding Chemical-Toxicity Information Resources in Support of Predictive Toxicology

Ann Richard
National Center for Computational Toxicology
US Environmental Protection Agency

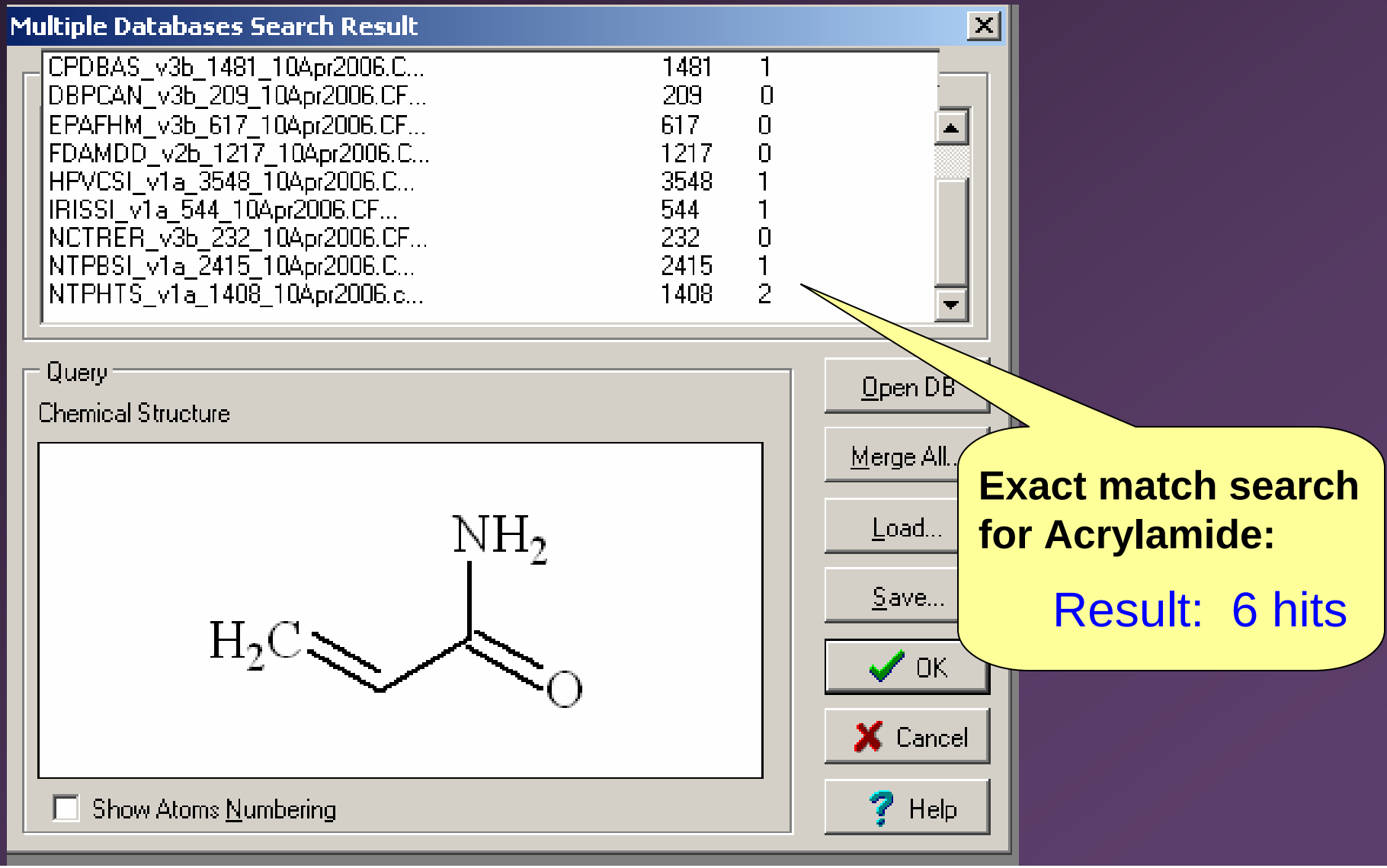


Past Lives:

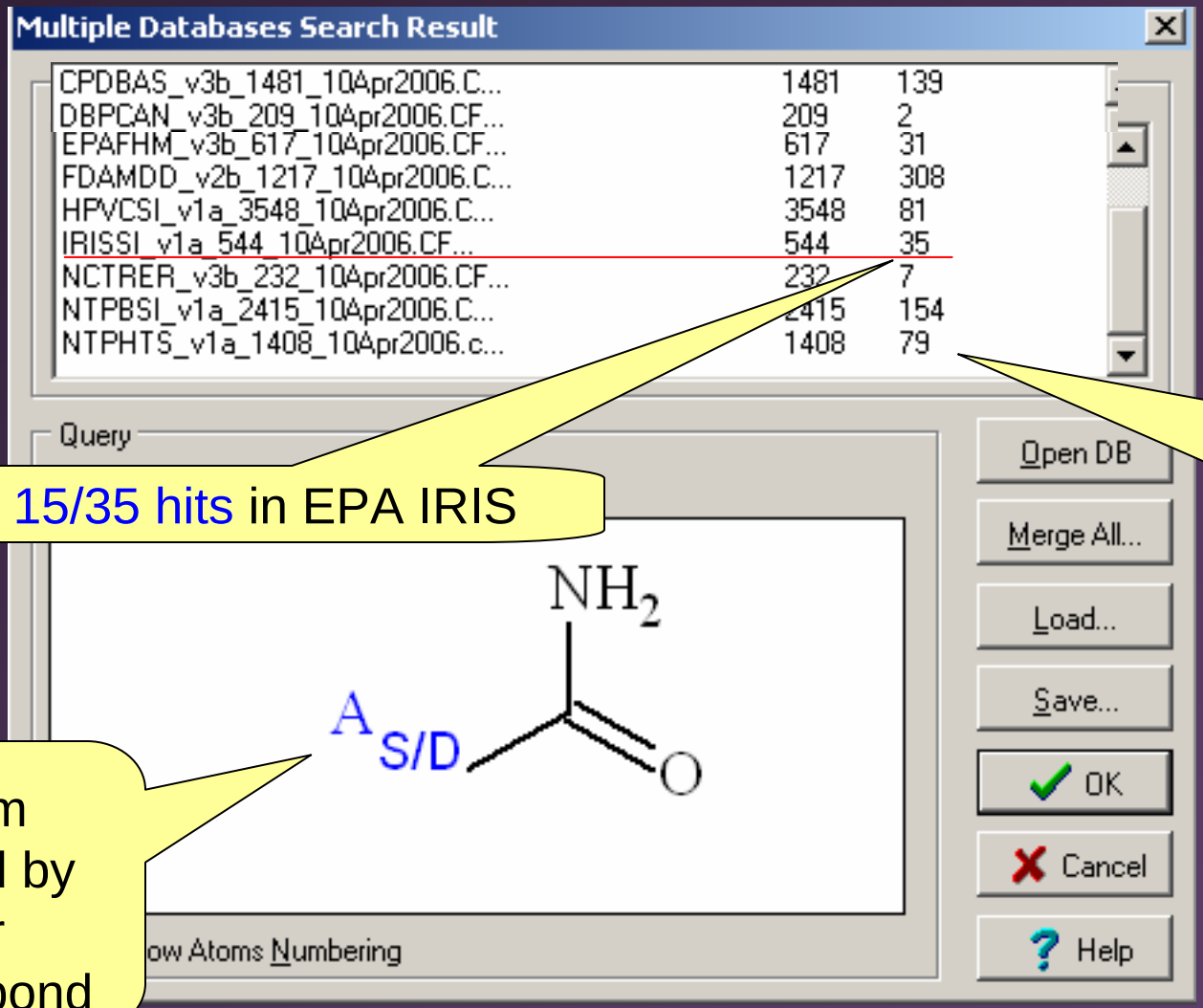
Molecular modeling, SAR, Chemoinformatics



Structure Searching Across 9 Diverse Toxicity Databases:



Generalized Sub-Structure Searching Across 9 Diverse Toxicity Databases:



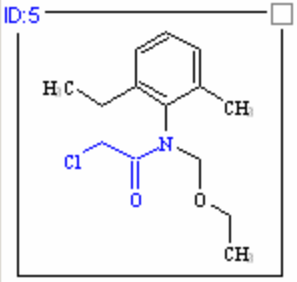
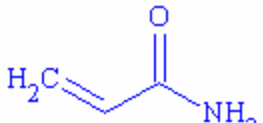
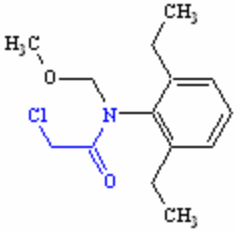
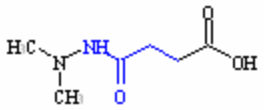
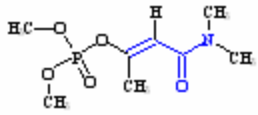
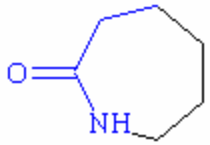
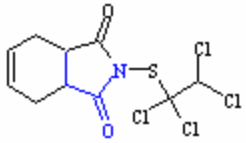
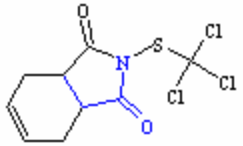
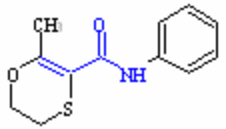
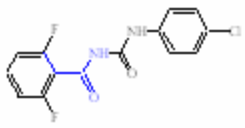
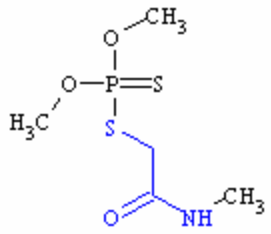
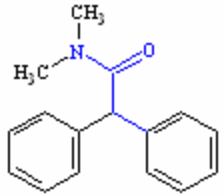
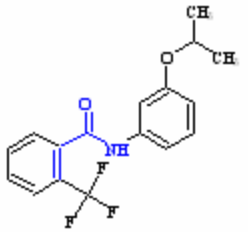
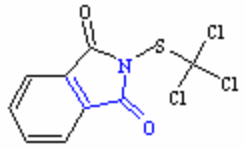
View 15/35 hits in EPA IRIS

Result: 836 hits

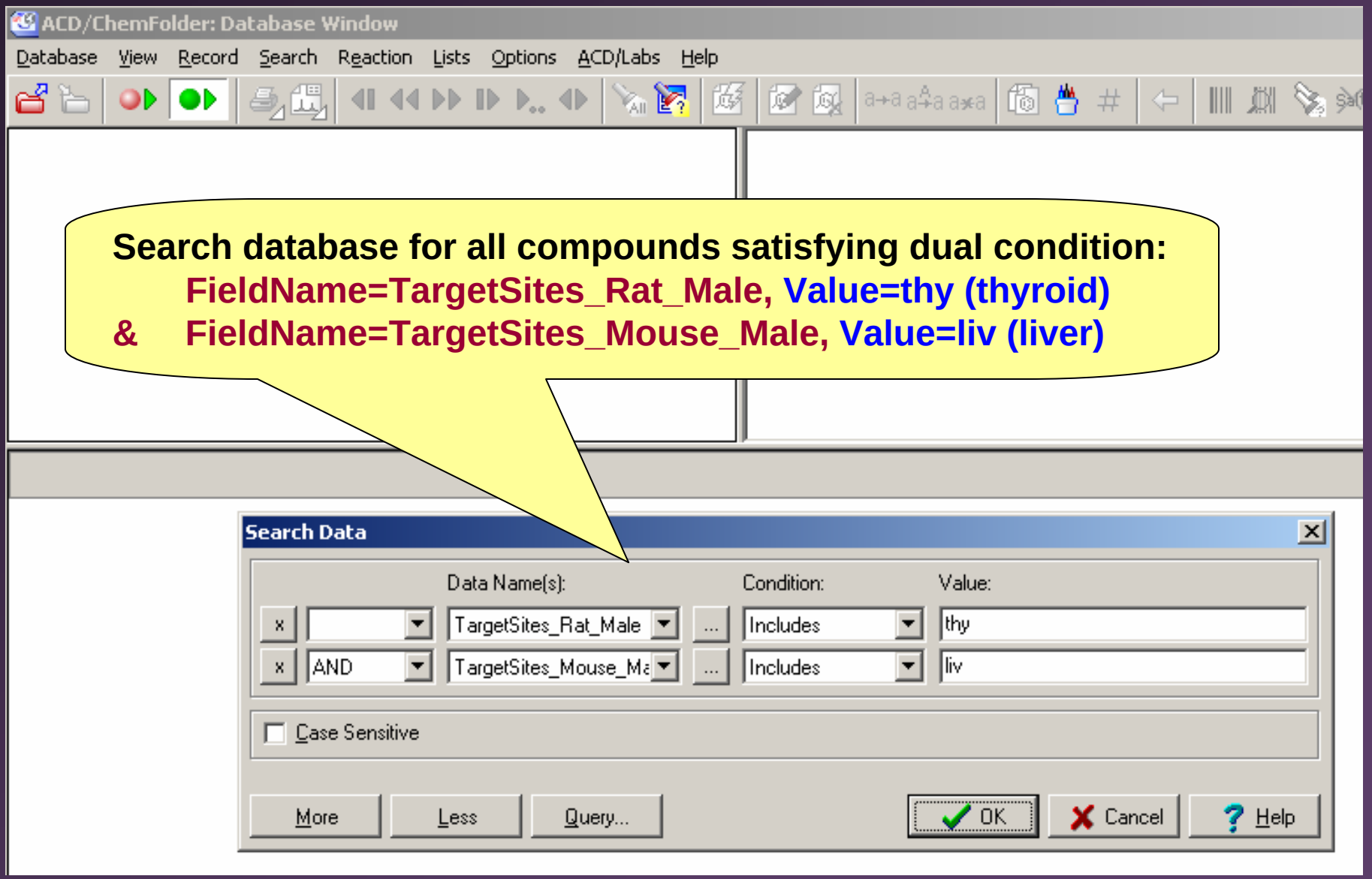
Any atom attached by single or double bond

Generalized Sub-Structure Searching Across 9 Diverse Toxicity Databases:

15/35 Hits in IRIS containing acrylamide-like moiety

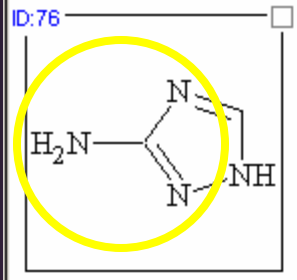
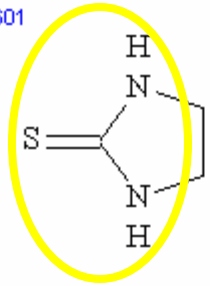
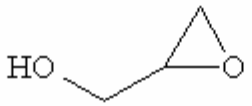
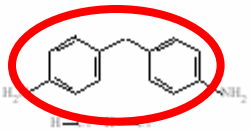
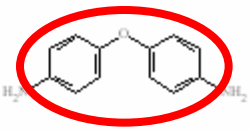
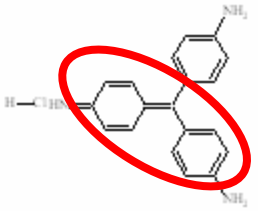
ID:5 	ID:12 	ID:16 	ID:17 	ID:58 
Acetochlor 34256-82-1	Acrylamide 79-06-1	Alachlor 15972-60-8	Daminozide 1596-84-5	Bentazon 25057-89-0
ID:71 	ID:99 	ID:100 	ID:101 	ID:108 
Bidrin 141-66-2	Caprolactam 105-60-2	Captafol 2425-06-1	Captan 133-06-2	Carboxin 5234-68-4
ID:206 	ID:210 	ID:229 	ID:266 	ID:268 
Diflubenzuron 35367-38-5	Dimethoate 60-51-5	Diphenamid 957-51-7	Flutolanil 66332-96-5	N-(Trichloromethylthio)phthali 133-07-3

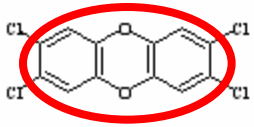
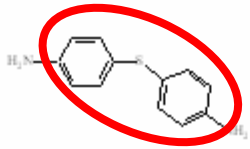
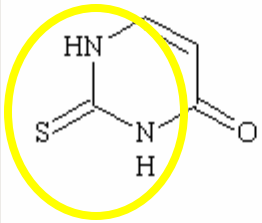
Relational Biological Content Searching: Carcinogenic Potency Database – All Species (CPDBAS_1481)



ACD/ChemFolder: Database Window - [D:\DSSTOX_MISC\ACD_CHEMFOLDERFILES_10APR2006\CPDBAS_V3B_1481_10APR2006.CFD]

Database View Record Search Reaction Lists Options ACD/Labs Help

ID:76 	ID:601 	ID:666 	ID:871 	ID:1094 	ID:1247 
76 CPDBAS_v3b_1481_10A	601 CPDBAS_v3b_1481_10	666 CPDBAS_v3b_1481_10	871 CPDBAS_v3b_1481_10	1094 CPDBAS_v3b_1481_1	1247 CPDBAS_v3b_1481_1
3-Aminotriazole	Ethylene thiourea (ETU)	Glycidol	4,4'-Methylenedianiline dihyd	4,4'-Oxydianiline	C.I. Basic red 9 monohydroc
negative	positive	positive	positive	positive	positive
thy	thy	ezy lgi mgl nrv per ski sto thy	liv thy	liv thy	ezy liv ski sub thy
pit thy	thy	cli hmo mgl nrv orc sto thy	thy	liv thy	ezy sub thy
liv	liv pit thy	hag liv lun ski sto	adr liv thy	hag liv	liv
liv	liv pit thy	hag mgl ski sub ute	hmo liv thy	hag liv thy	adr liv

ID:1315 	ID:1344 	ID:1347 
1315 CPDBAS_v3b_1481_1	1344 CPDBAS_v3b_1481_1	1347 CPDBAS_v3b_1481_1
2,3,7,8-Tetrachlorodibenzo-p	4,4'-Thiodianiline	Thiouracil
negative	positive	thy
orc thy	ezy lgi liv thy	thy
liv lun	ezy thy ute	liv
liv	liv thy	liv
liv thy	liv thy	

Search Results: 9 hits / 1481 total

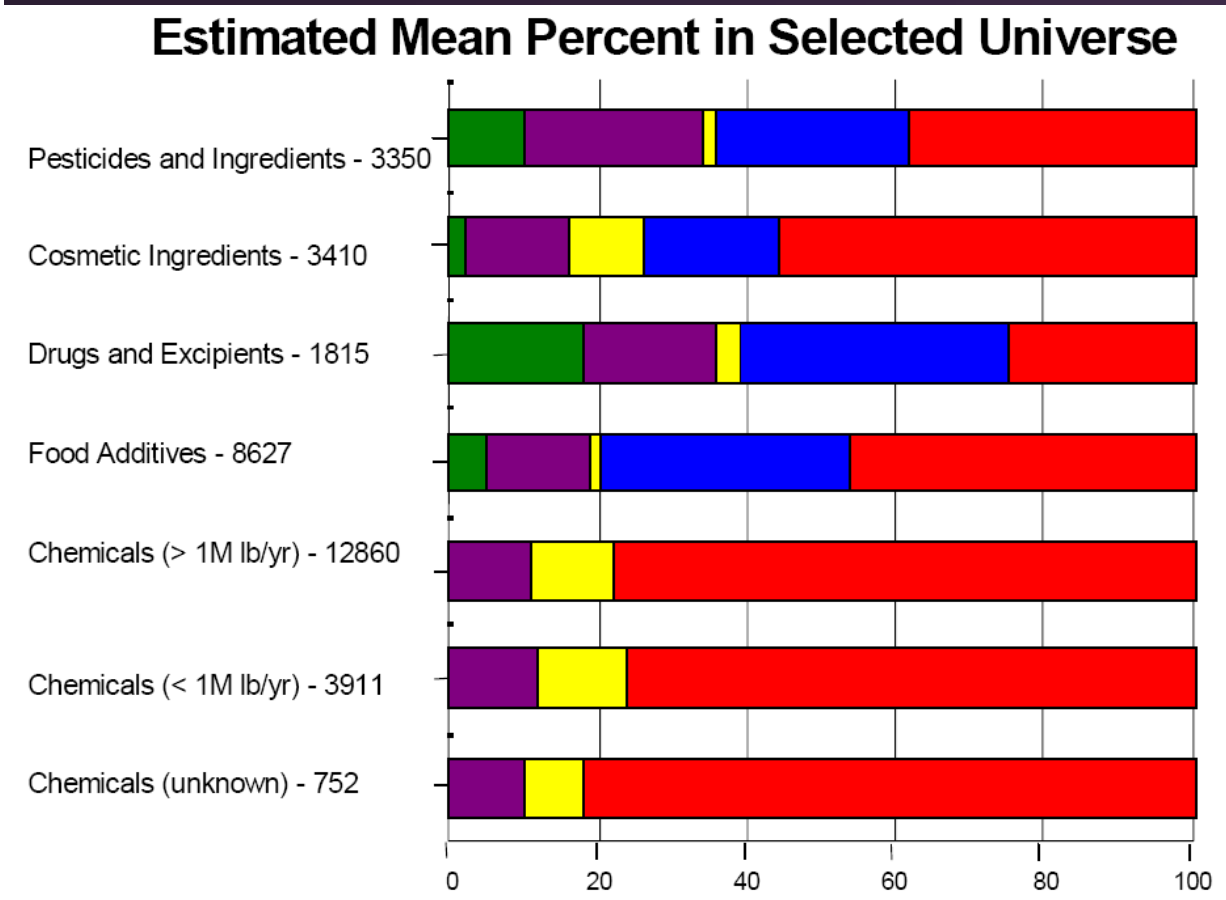
TargetSites_Rat_Male = thyroid

& TargetSites_Mouse_Male = liver

Part I.

The Problem

Environmental Chemicals: Toxicity Assessment Data Gaps



Complete Hazard Assessment Possible

Minimal Information Available

No Toxicity Information Available

Partial Hazard Assessment Possible

Some Toxicity Information Available

Strategies for Closing the Chemical Data Gap

by John S. Applegate and Katherine Baer

A Center for Progressive Reform Publication *April 2006*

EPA Problems:

- Large lists of chemicals to evaluate
- Many toxicity endpoints to assess
- Lack of sufficient and relevant data

Need to prioritize and focus limited resources on chemicals and problem areas with potential for greatest health & environmental impact

Endocrine Disruption Testing Program

EPA ORD Human Health Risk Assessment Strategy Document

Cross-Chemical Extrapolation

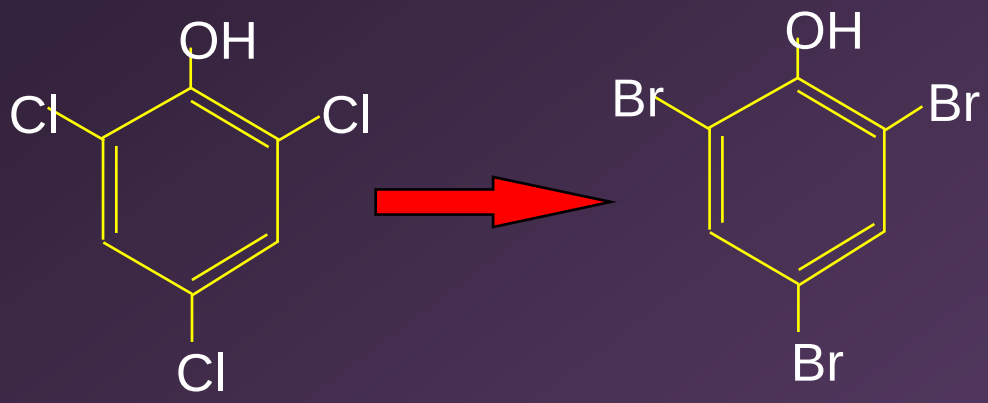
A major need in risk assessment is to *improve methods for addressing the large numbers of chemicals for which little or no toxicity data are available*.

…

This extrapolation procedure assumes that the behavior of chemicals in biological systems ... can be inferred or projected based on *analogies* and correlations to structural, physicochemical, or electronic aspects of the parent compound or its metabolites.

The underlying bases of these analogies and correlations is that *these commonalities will cause the chemicals to behave in a parallel manner in biological systems, i.e., they will possess a common mode-of-action.*

EPA Regulatory Action on 2,4,6-tribromophenol



2,4,6-trichlorophenol

genotoxic
forms free radicals
DNA reactive
rodent carcinogen

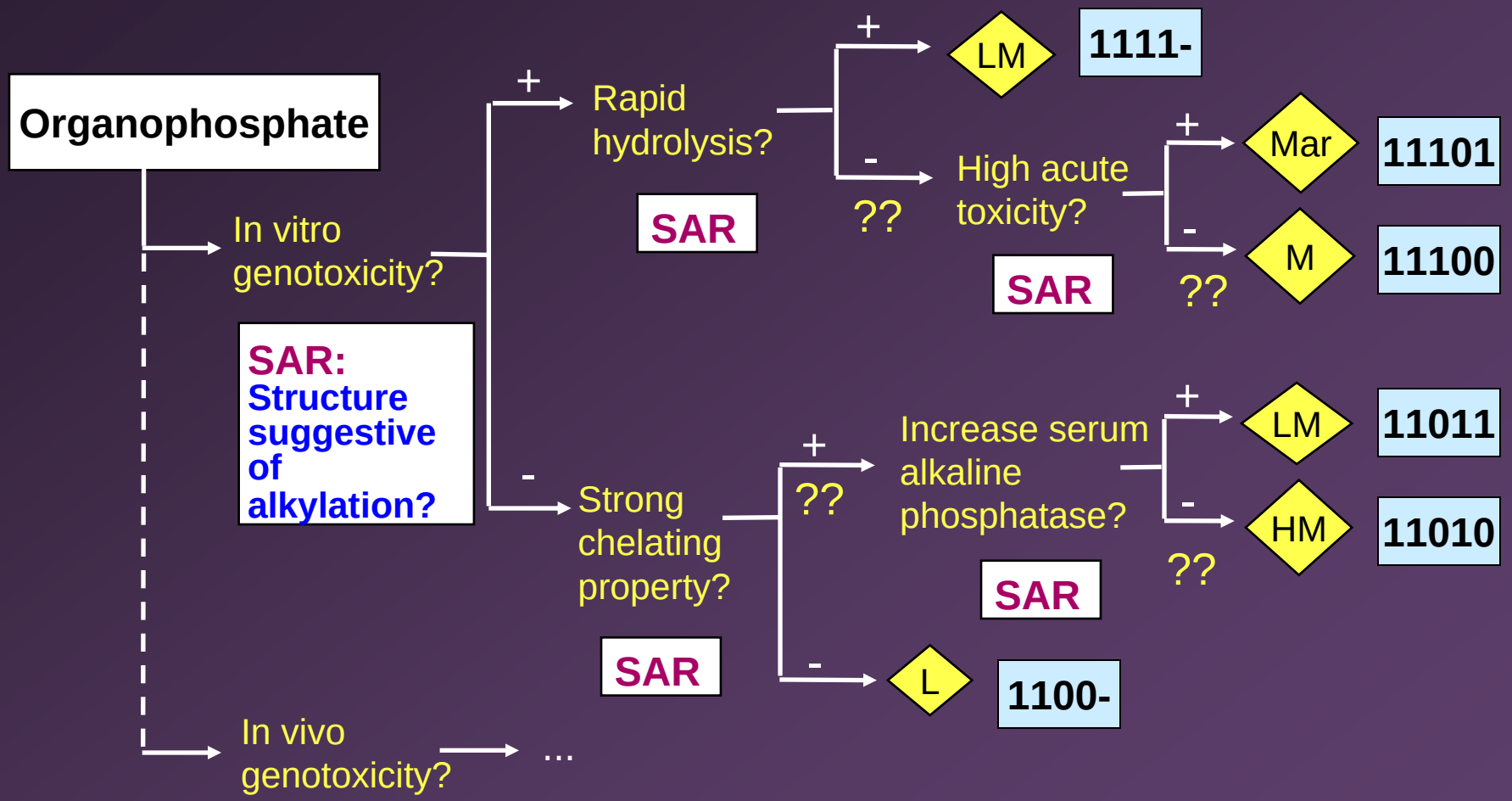
2,4,6-tribromophenol

genotoxic
forms free radicals
DNA reactive
???

Chemical structural analogy
Biological mechanism analogy

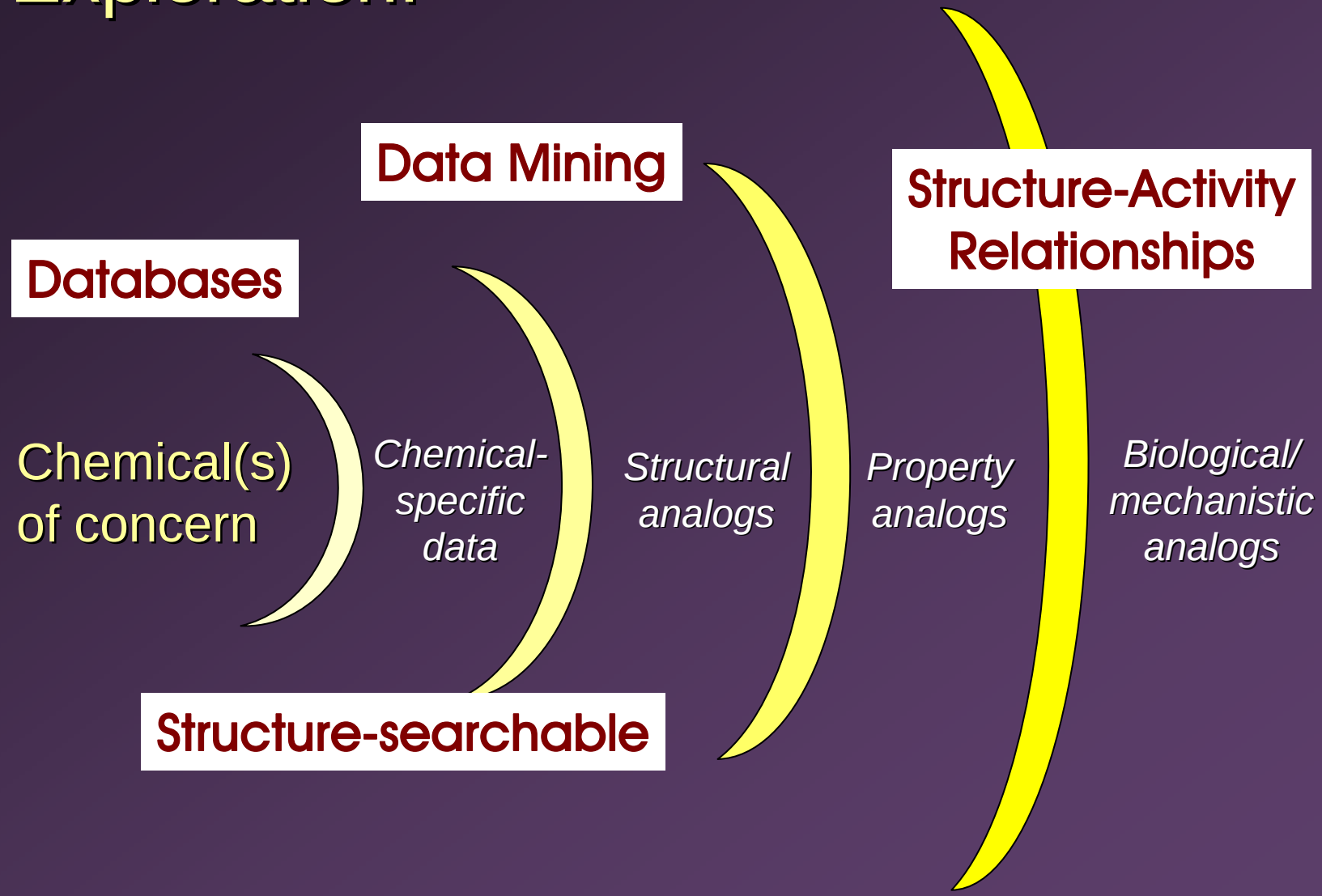
Combining SAR and Biofunctional Information

Predicting Carcinogenicity of Organophosphates



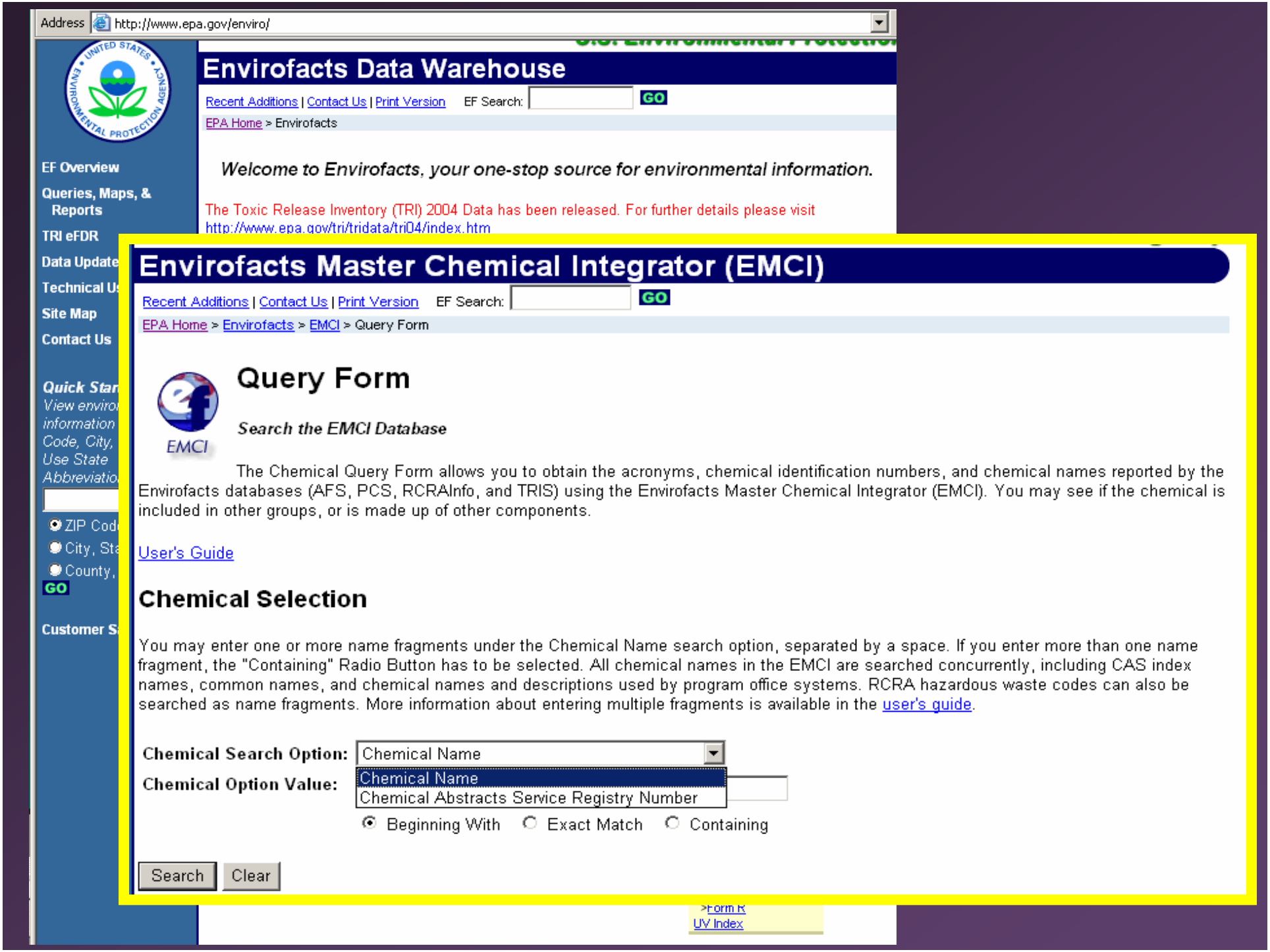
Woo et al. (1996) Environ. Carc. & Ecotox. Revs., C14:1-42

Chemistry-based Data Mining & Exploration:



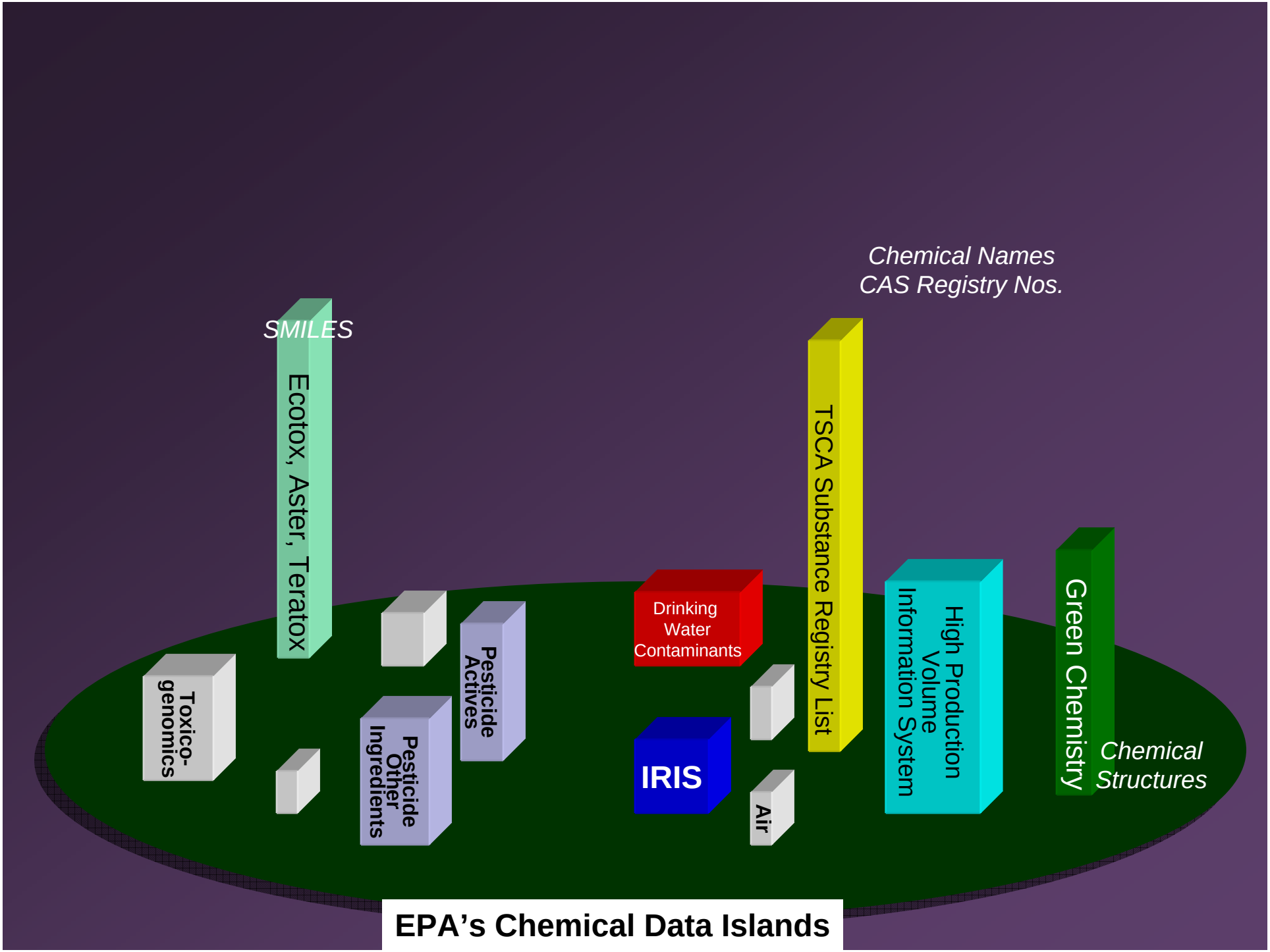
Part II.

Data-Mining of Public Toxicity Databases

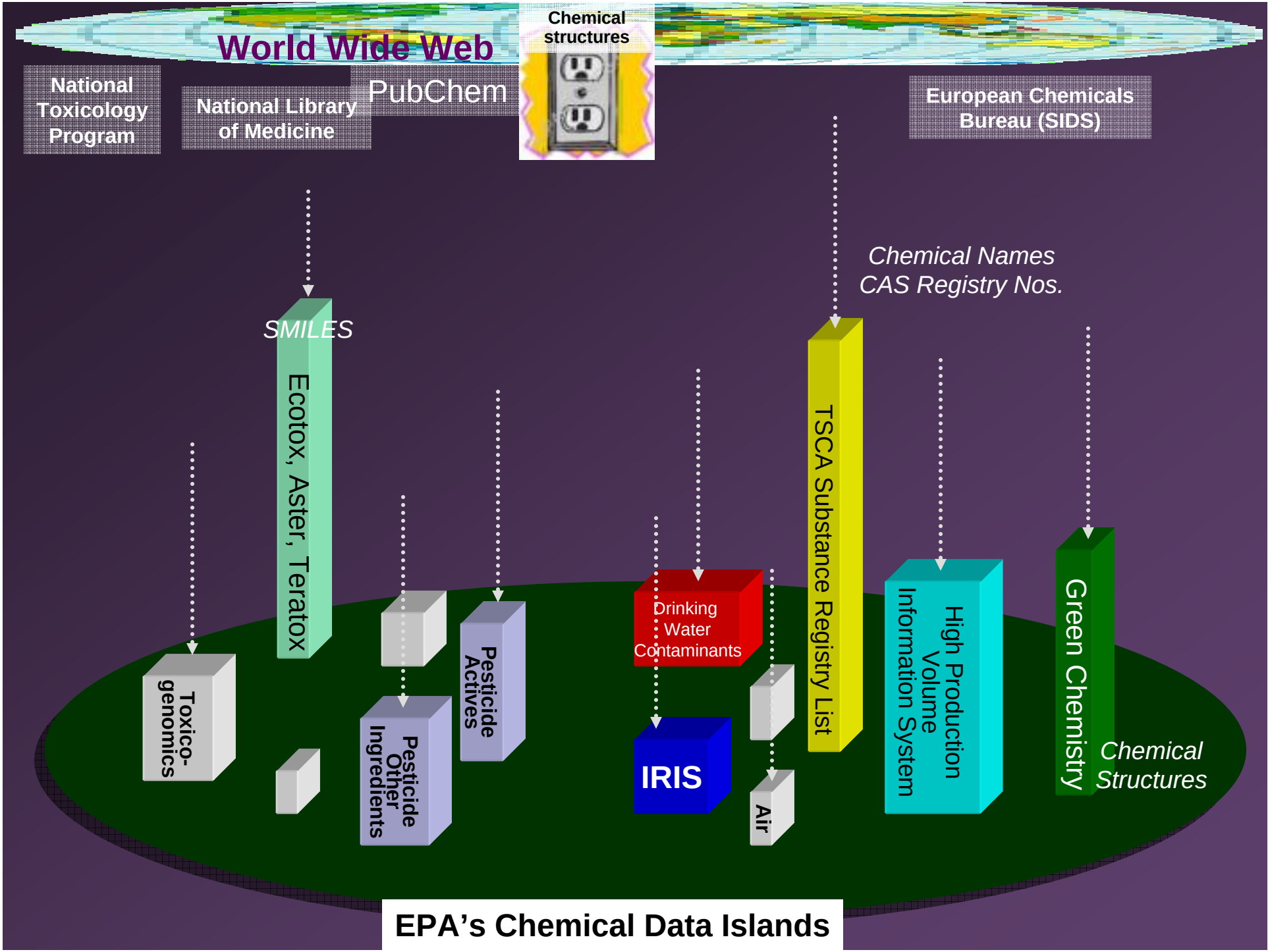

Address http://www.epa.gov/enviro/
Envirofacts Data Warehouse
Recent Additions | Contact Us | Print Version EF Search: GO
EPA Home > Envirofacts
Welcome to Envirofacts, your one-stop source for environmental information.
The Toxic Release Inventory (TRI) 2004 Data has been released. For further details please visit
http://www.epa.gov/tri/tridata/tri04/index.htm
EF Overview
Queries, Maps, & Reports
TRI eFDR
Data Update
Technical U
Site Map
Contact Us
Quick Star
View enviro
information
Code, City,
Use State
Abbreviatio
ZIP Cod
City, Sta
County,
GO
Customer S
Envirofacts Master Chemical Integrator (EMCI)
Recent Additions | Contact Us | Print Version EF Search: GO
EPA Home > Envirofacts > EMCI > Query Form
EMCI
Query Form
Search the EMCI Database
The Chemical Query Form allows you to obtain the acronyms, chemical identification numbers, and chemical names reported by the Envirofacts databases (AFS, PCS, RCRAInfo, and TRIS) using the Envirofacts Master Chemical Integrator (EMCI). You may see if the chemical is included in other groups, or is made up of other components.
User's Guide
Chemical Selection
You may enter one or more name fragments under the Chemical Name search option, separated by a space. If you enter more than one name fragment, the "Containing" Radio Button has to be selected. All chemical names in the EMCI are searched concurrently, including CAS index names, common names, and chemical names and descriptions used by program office systems. RCRA hazardous waste codes can also be searched as name fragments. More information about entering multiple fragments is available in the user's guide.
Chemical Search Option: Chemical Name
Chemical Option Value:
Chemical Name
Chemical Abstracts Service Registry Number
Beginning With Exact Match Containing
Search Clear
>Form R
UV Index

Problems across EPA:

- Little or no chemical structure annotation
- No structure searching capabilities (Internet or desktop)
- No coordination of chemical structure information across EPA databases
- No standard quality review procedures for chemical information



EPA's Chemical Data Islands


World Wide Web
Chemical structures
National Toxicology Program
National Library of Medicine
PubChem
European Chemicals Bureau (SIDS)
Chemical Names
CAS Registry Nos.
SMILES
Ecotox, Aster, Teratox
TSCA Substance Registry List
High Production Volume Information System
Green Chemistry
Chemical Structures
Toxico-genomics
Pesticide Actives
Pesticide Other Ingredients
Drinking Water Contaminants
IRIS
Air
EPA's Chemical Data Islands

National Toxicology Program

Department of Health and Human Services

Home | Testing Information | Study Results & Research Projects | Public Health | About the NTP | Help

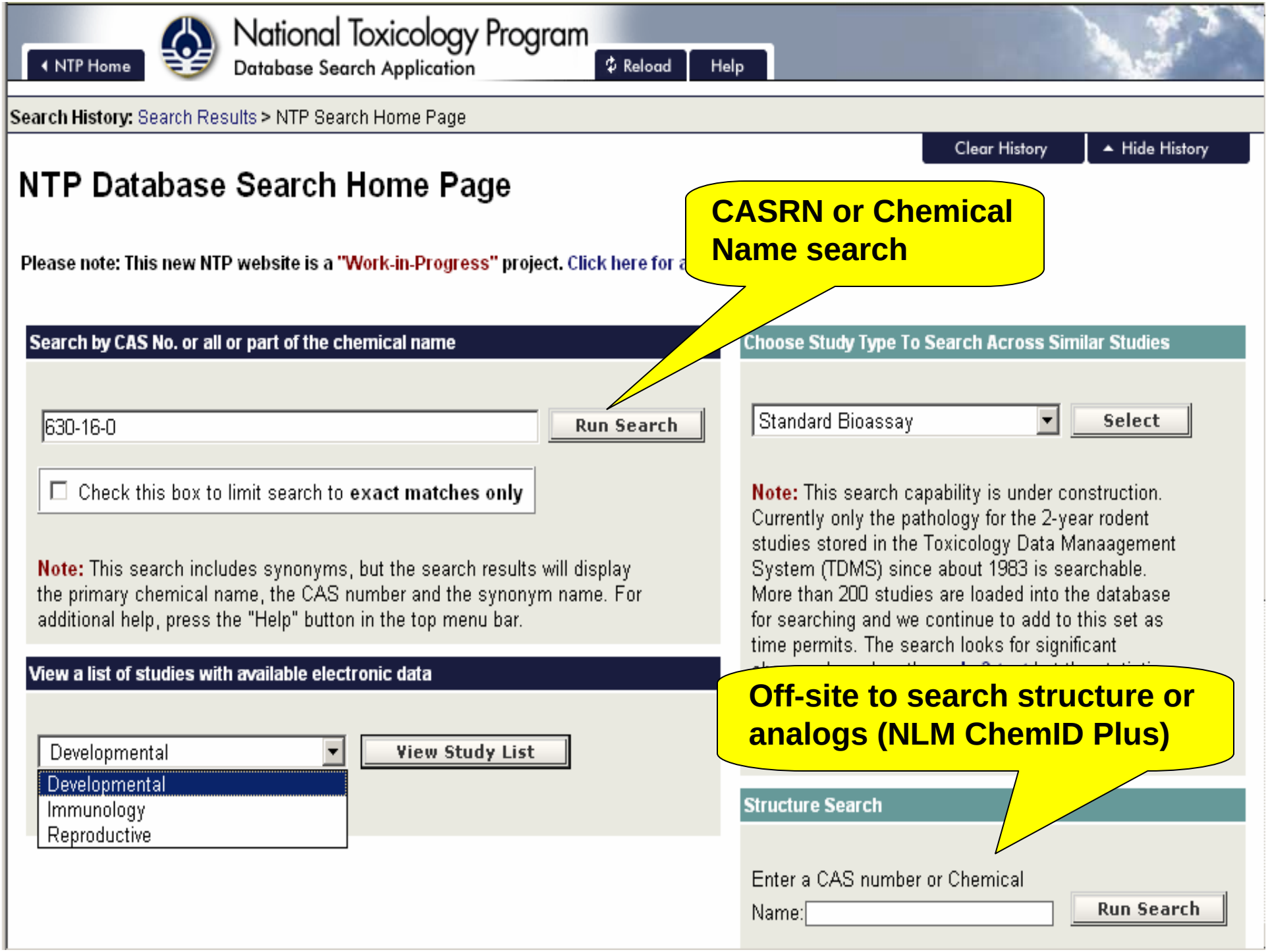
http://ntp.niehs.nih.gov/

- NTP Study Reports
- Study Data Searches
- Pathology Tables for Peer Review
- Summaries & Associations of Study Results
- NTP Center for Phototoxicology
- Center for Rodent Genetics
- New Areas of Research

Reports: NTP is converting study reports into an electronic format which can be accessed from the website. These reports are made available as soon as they have been converted.

Data Searches: The NTP has been loading study information into databases and has developed applications to access this data from the web. There are two types of data mining searches:

- All types of data - search provides a way to find the various types of studies conducted on a test agent and has options to mine that data if it is available in electronic format.
- Bioassay pathology data mining search provides a way to access the pathology databases. It is also possible to search the historical control database and to view


NTP Home
National Toxicology Program
Database Search Application
Reload
Help
Search History: Search Results > NTP Search Home Page
Clear History
Hide History
NTP Database Search Home Page
Please note: This new NTP website is a "Work-in-Progress" project. Click here for a
CASRN or Chemical Name search
Search by CAS No. or all or part of the chemical name
630-16-0
Run Search
Check this box to limit search to exact matches only
Note: This search includes synonyms, but the search results will display the primary chemical name, the CAS number and the synonym name. For additional help, press the "Help" button in the top menu bar.
View a list of studies with available electronic data
Developmental
Developmental
Immunology
Reproductive
View Study List
Choose Study Type To Search Across Similar Studies
Standard Bioassay
Select
Note: This search capability is under construction. Currently only the pathology for the 2-year rodent studies stored in the Toxicology Data Manaagement System (TDMS) since about 1983 is searchable. More than 200 studies are loaded into the database for searching and we continue to add to this set as time permits. The search looks for significant
Off-site to search structure or analogs (NLM ChemID Plus)
Structure Search
Enter a CAS number or Chemical Name:
Run Search

NTP Studies on 1,1,1,2-Tetrabromoethane

Table Instructions and Notes:

5 study categories

- Choose study type to view data for **1,1,1,2-Tetrabromoethane**
- Not all agents have been studied in every study type. If there are no electronically availab[illegible]a on **1,1,1,2-Tetrabromoethane**, you may contact the NTP Central Files (cdm@niehs.nih.gov) to request available data from completed st[illegible]es or the status for ongoing studies

Standard Toxicology & Carcinogenesis Studies	Reproductive Studies	Developmental Studies	Immunology Studies	Genetic Toxicity Studies
Description of standard protocols Study C91016 **Status:** Rpt Complete **Length:** 3-Week **Route:** Gavage **Rats:** FISCHER 344 No data available online. View Study Abstract	No Reproductive Studies Available for this Chemical	No Developmental Studies Available for this Chemical	No Immunology Studies Available for this Chemical	Description of standard protocols In Vitro Study Data • Salmonella

View study results:
Genetic toxicity studies: Salmonella

National Toxicology Program
Database Search Application

NTP Home | New Search | Return to Study List | Reload | Help

Search History: Search Results > NTP Search Home Page > Search Results > NTP Studies on 1,1,1,2-Tetrabromoethane > Salmonella S

Salmonella Study Overview

Current Search Criteria	
Chemical Name:	1,1,1,2-Tetrabromoethane
CAS Number:	630-16-0
Study Type:	Salmonella

Table Instructions and Notes:

Click on the study number to view a summary of the results.

Standard NTP Protocol

Study Summary Call

Study ID	Study Result	Year Completed
A87711	Weak Positive	1993

View study details

Chemical Name:	1,1,1,2-Tetrabromoethane
CAS Number:	630-16-0
Study Type:	Salmonella
Study ID:	A87711

Summary Table

Study Summary Call

Study ID	Overall Result	Year Completed
A87711	Weak Positive	1993

No "Endpoint" Summary Calls

Table Instructions and Notes:

Click 'View Detailed Data' to proceed to the study data.

Options	Strain	S9 Activation	S9 Species	Concentration
View Detailed Data	TA100	Yes	Hamster	10% HLI
	TA100	Yes	Rat	10% RLI
	TA100	Yes	Hamster	30% HLI
	TA100	Yes	Rat	30% RLI
	TA100	No	-	-
	TA1535	Yes	Hamster	10% HLI
	TA1535	Yes	Rat	10% RLI
	TA1535	Yes	Hamster	30% HLI
	TA1535	Yes	Rat	30% RLI

View detailed data

Strain: TA1538

Dose	30% RLI (Negative)		30% HLI (Negative)	
ug/Plate	Mean	Std. Error	Mean	Std. Error
Vehicle C				10
				1
				30
				20
				20
				30
Positive C				30

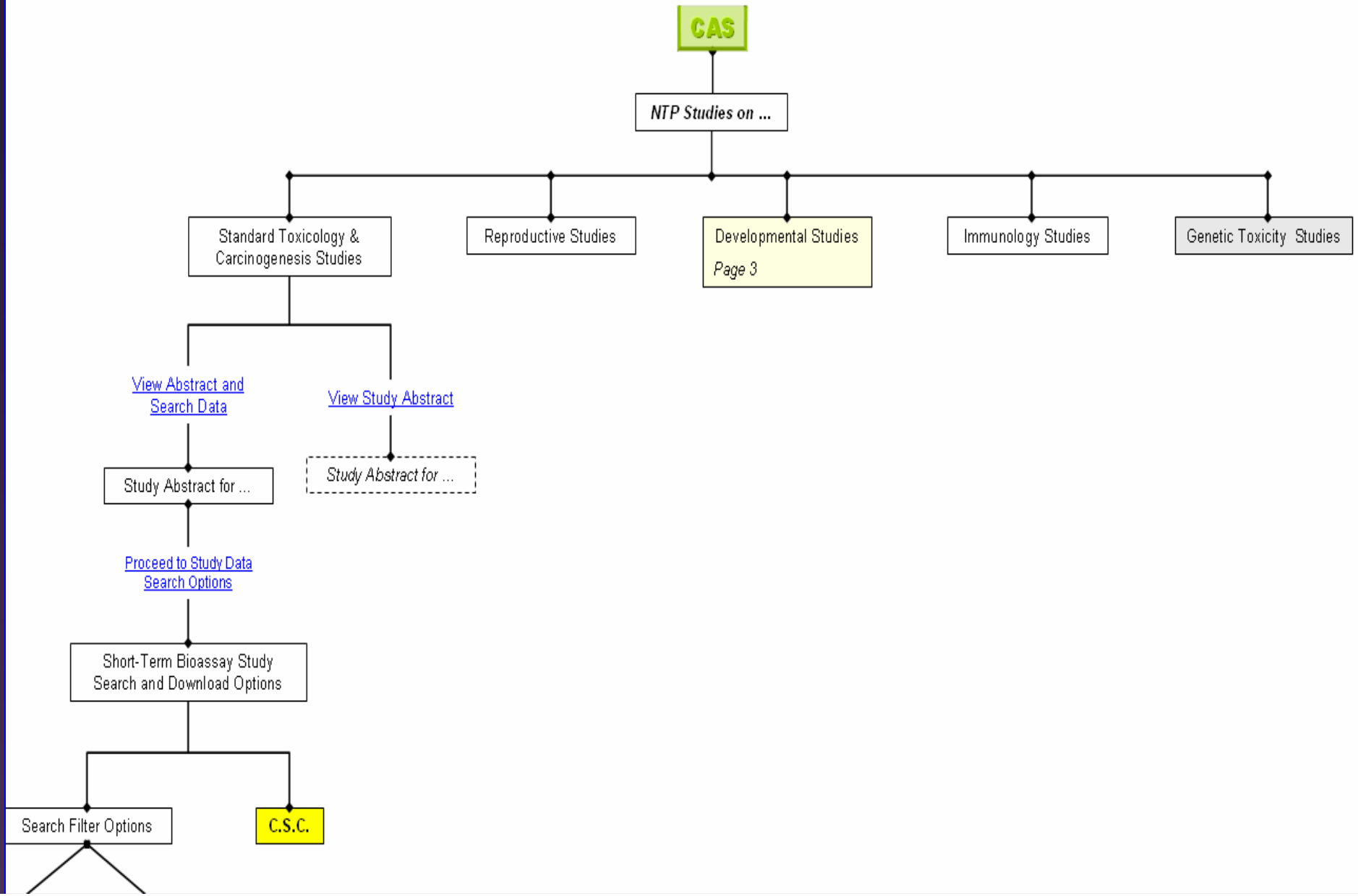
Individual experiment results

- Cannot download entire list of NTP chemicals and test summary data
- Cannot structure or substructure-search database
- Cannot download subsets of data:
  - list of TA98 pos data
  - list of all thyroid tumor carcinogens
- Cannot ask relational questions of data:
  - what chemicals are TA100 neg + TA98 pos?
  - list all chemicals with positive rat liver tumor findings in cancer bioassay that are also non-mutagenic

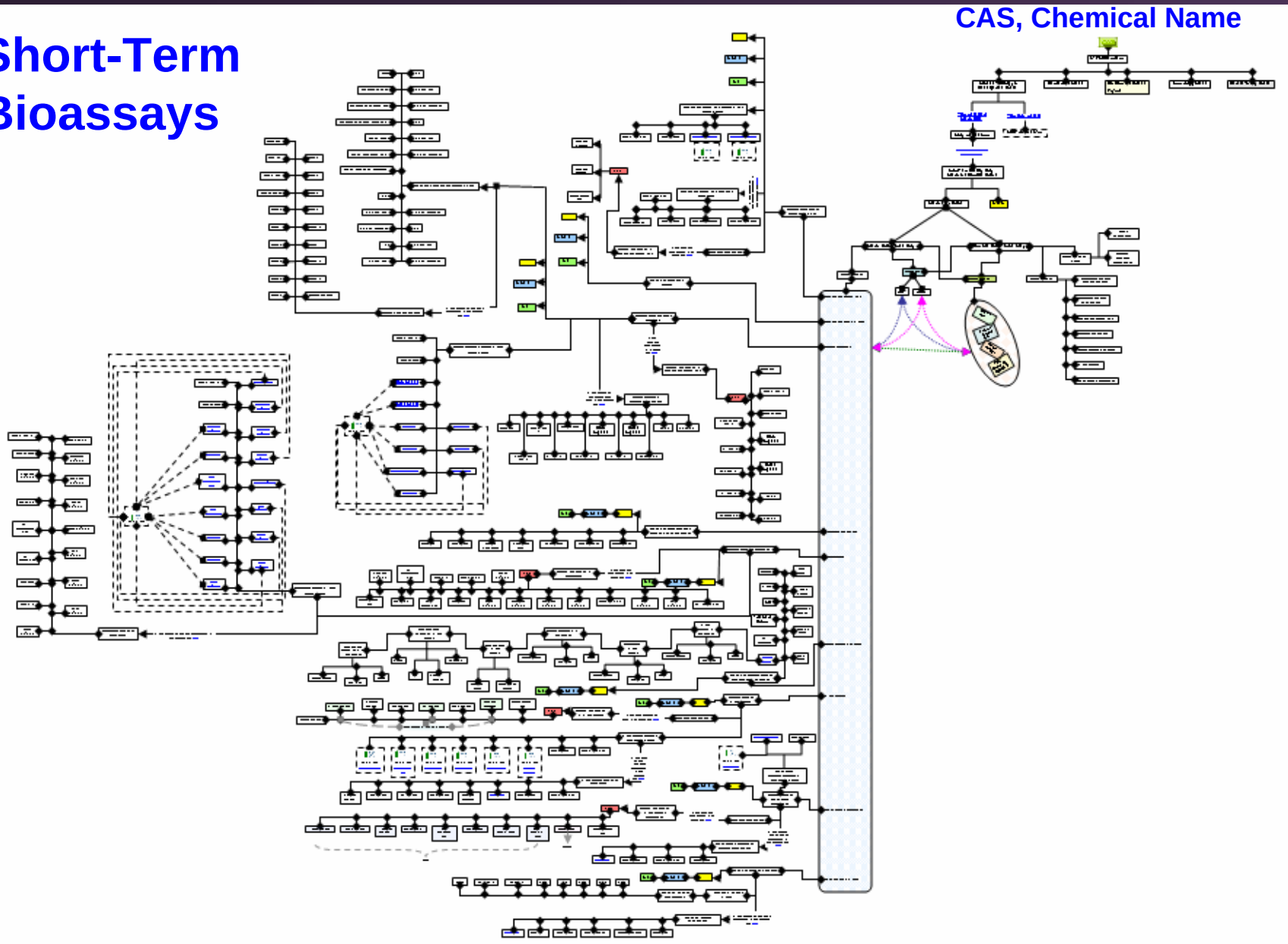
Strain: TA97

Dose	10 (Ne								
ug/Plate	Mean								
Vehicle Control	179								30
100	168								30
200	192								30
400	T								9
600	T								30
800	T	0	T	0	T	0	T	0	
Positive Control	734	28	376	14.40	1621	36.40	678	19.90	

NTP Database Site Map



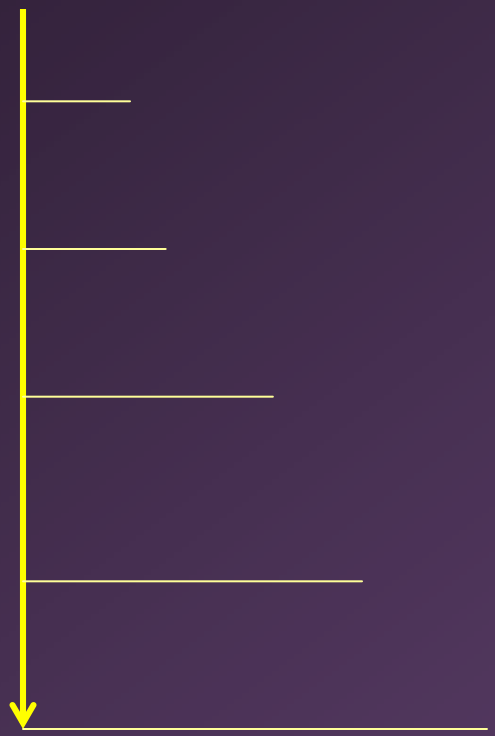
NTP Database Site Map

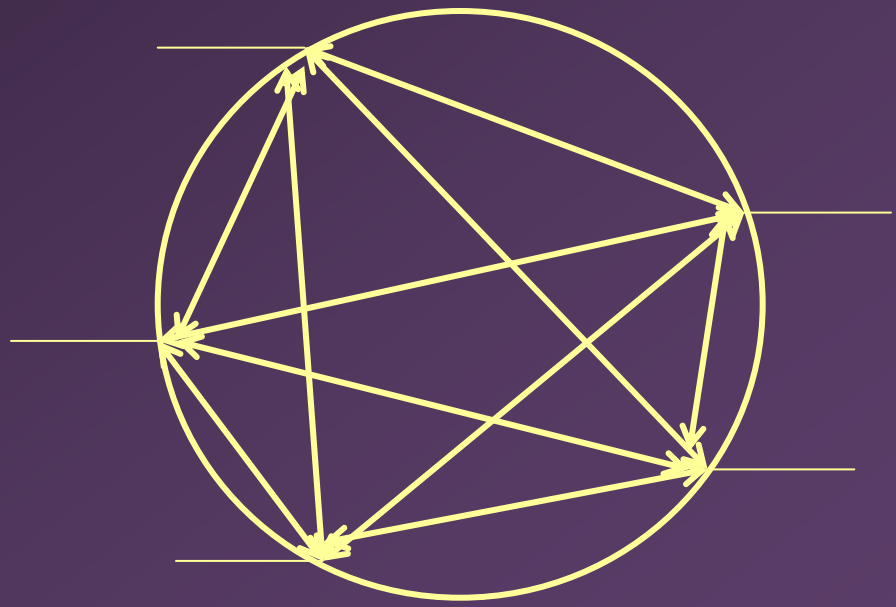


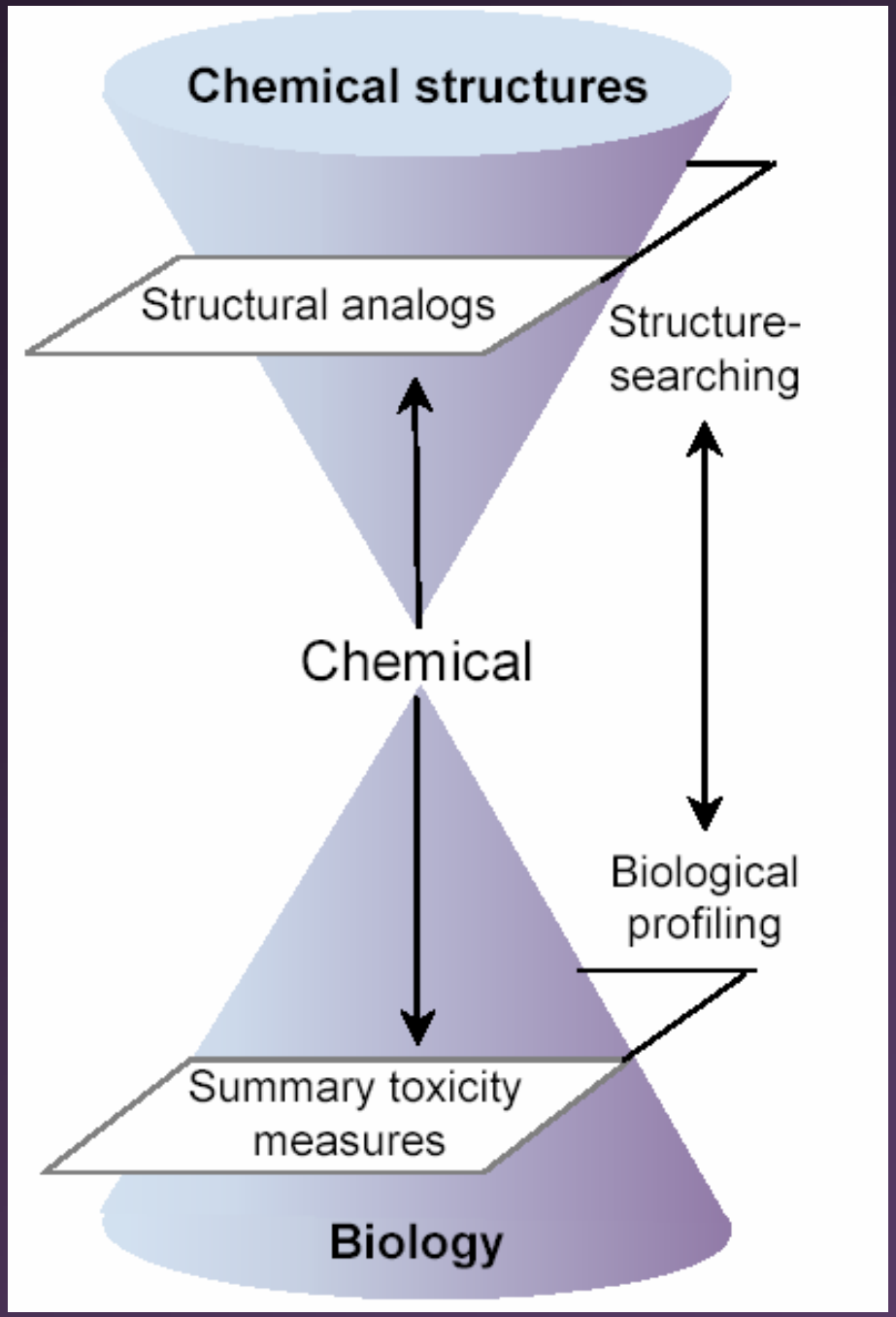
Archival

vs.

Relational





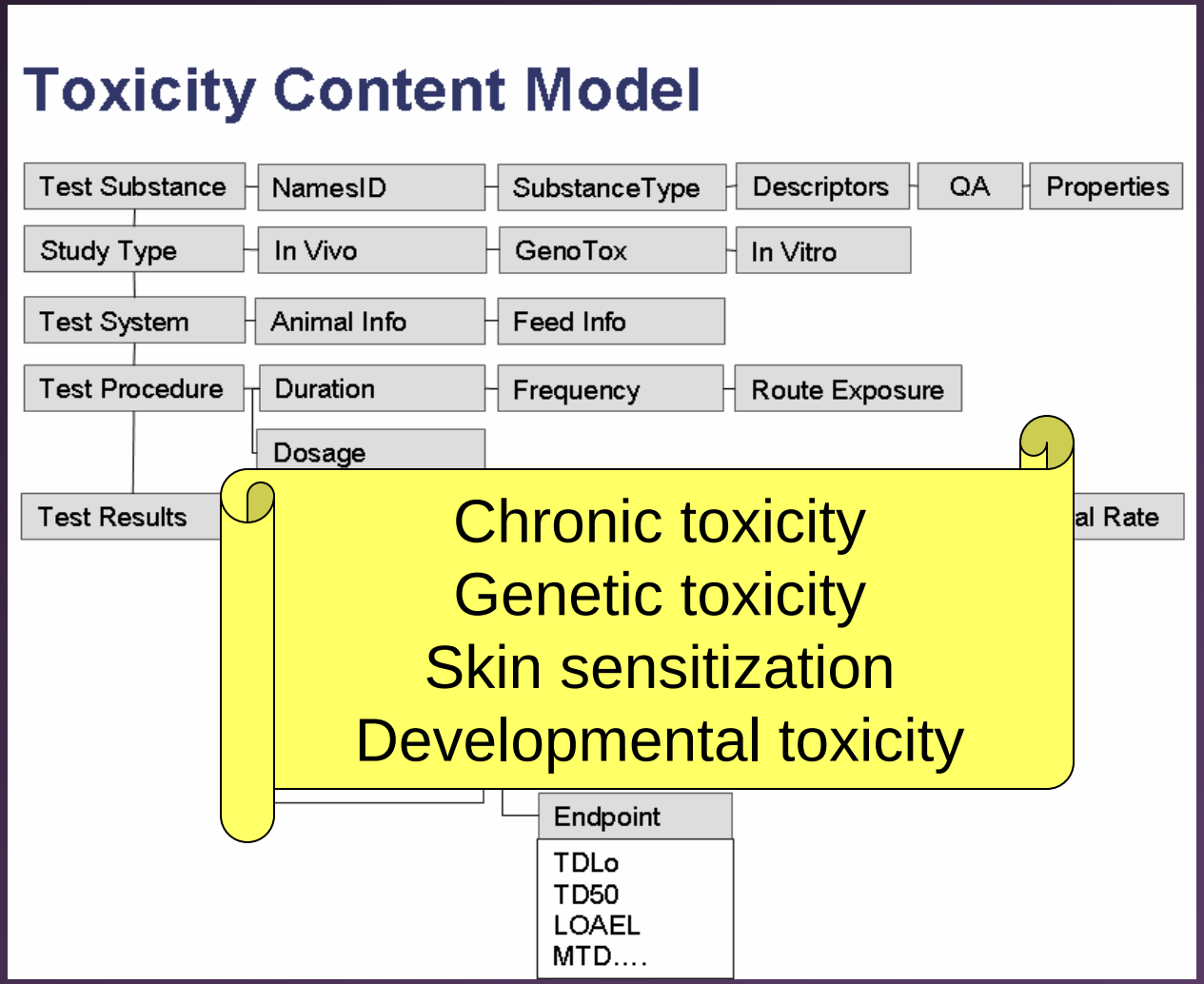


Yang et al (2006) Landscape of current toxicity databases and database standards. *Curr Opinion Drug Discov Develop* 9(1),124-133.

Yang et al (2006) The art of data mining the minefields of toxicity databases to link chemistry to biology. *Curr Comput-Aided Drug Design*, 2(2), 135-150.

ToxML Public Data Model Schemas

LIST Consortium



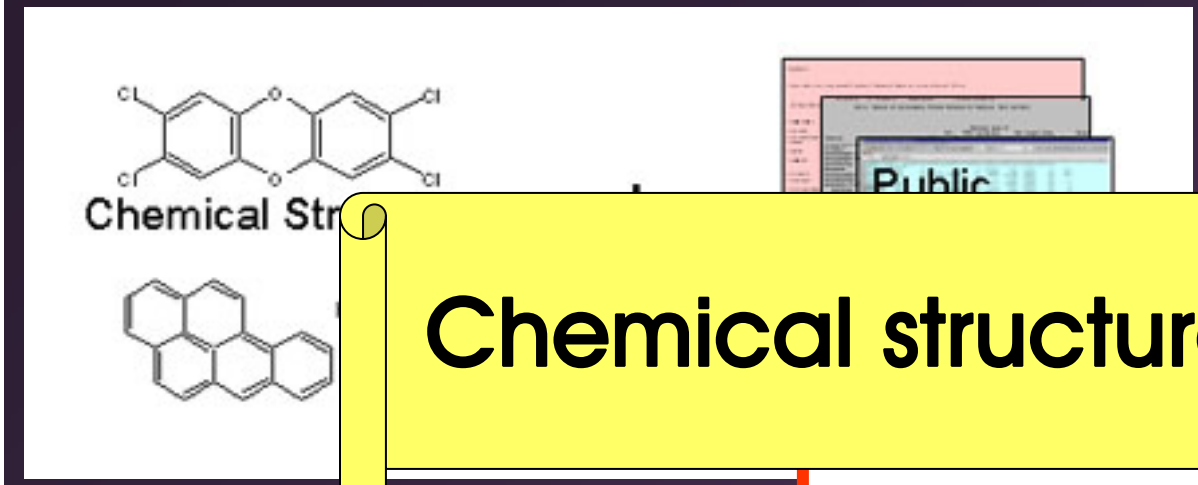
- Standard data forms
- Controlled vocabulary
- Hierarchical field structure
- Toxicity domain expert involvement
- XML format
- Top-level chemical indexing

FDA CDER/CFSAN ToxML Database Collaborations

LeadScope

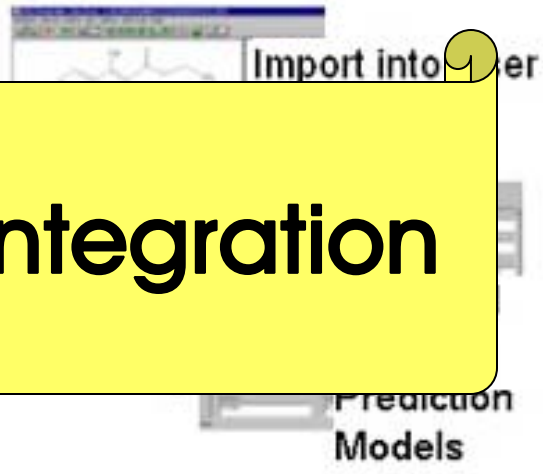
Part III.

The DSSTox Project



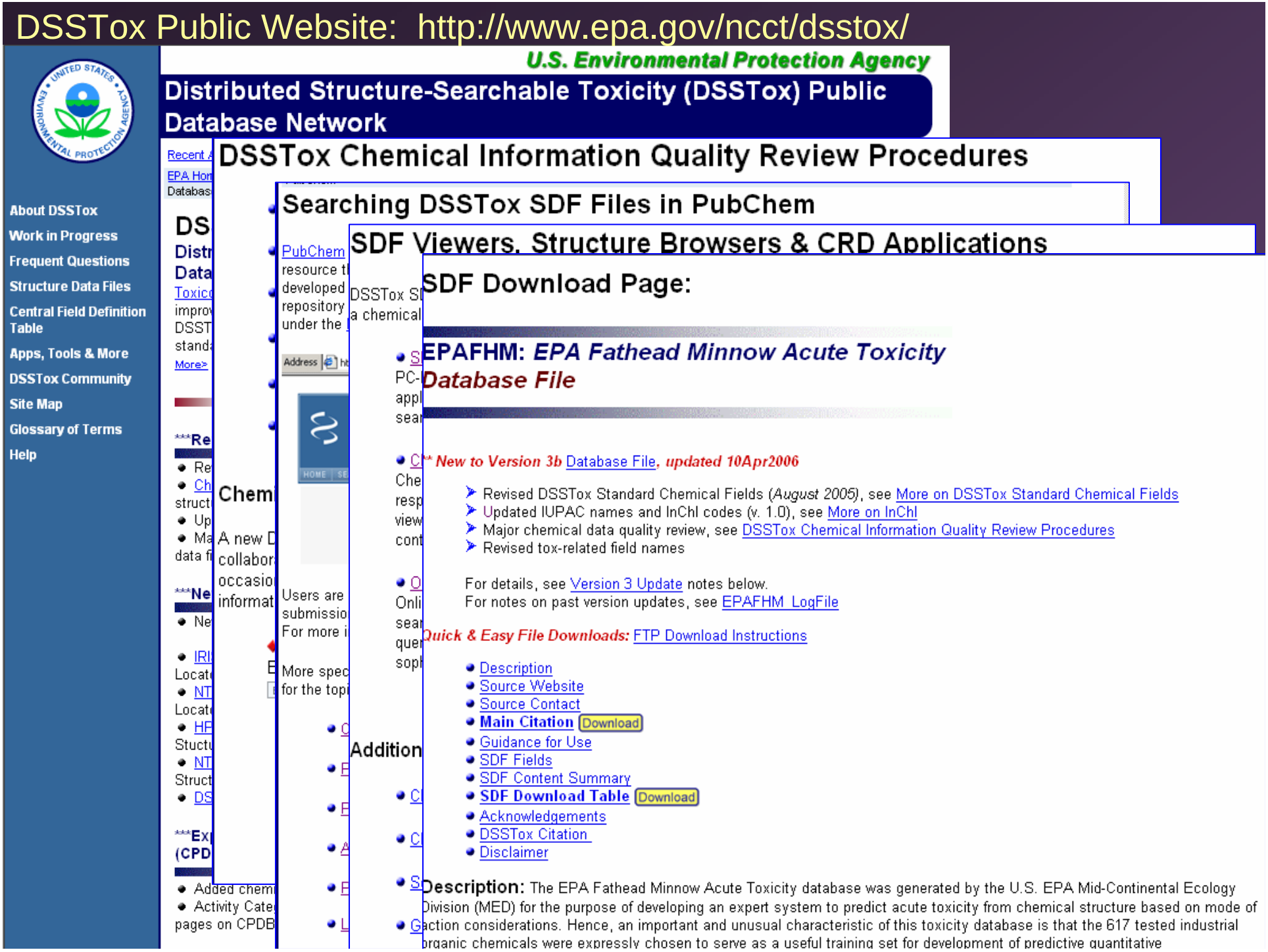
Chemical structure-annotation

Distributed
Structure-**S**earchable
Toxicity
Public
Database
Network

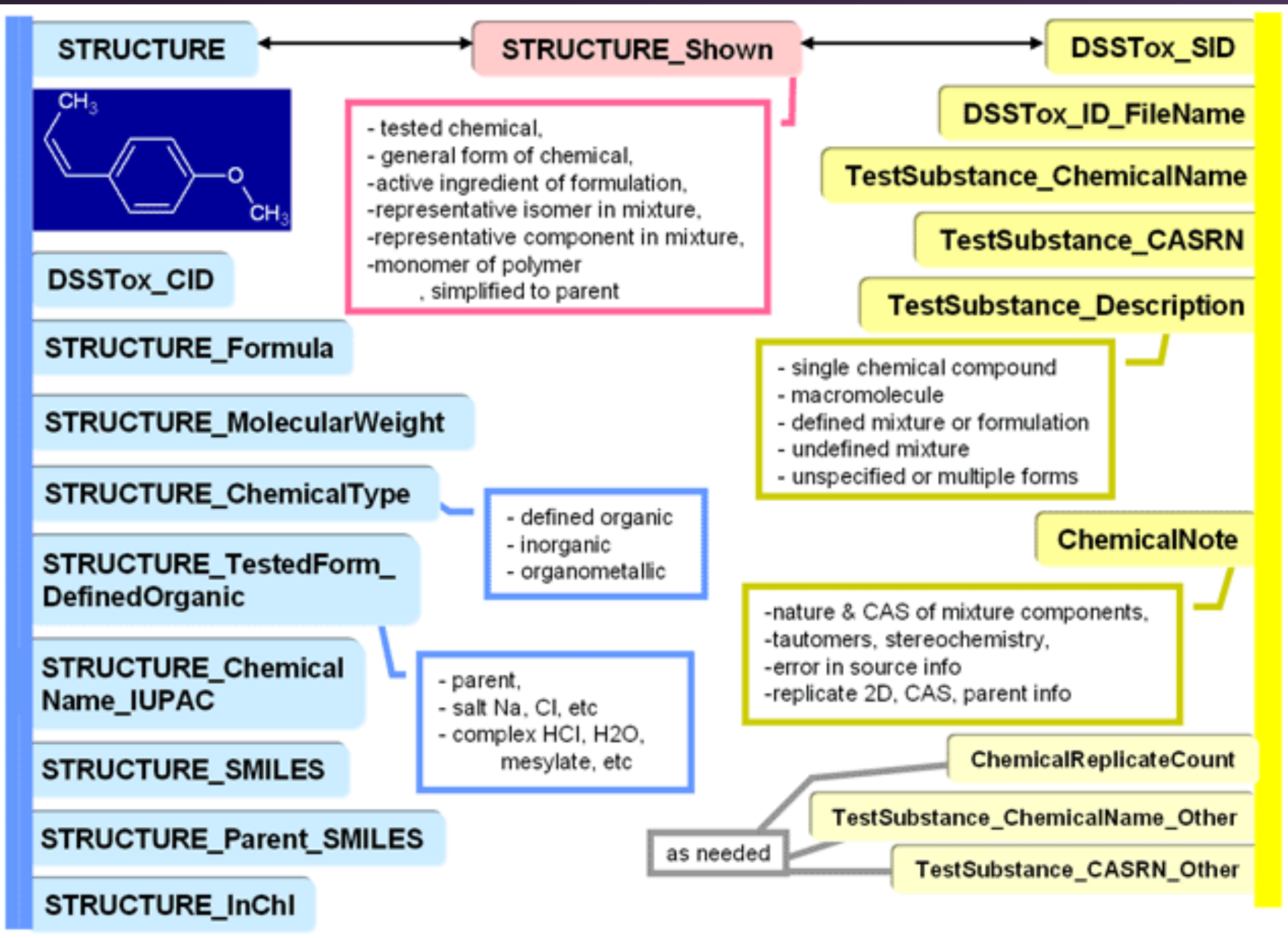


Data standards and integration

DSSTox Public Website: http://www.epa.gov/ncct/dsstox/



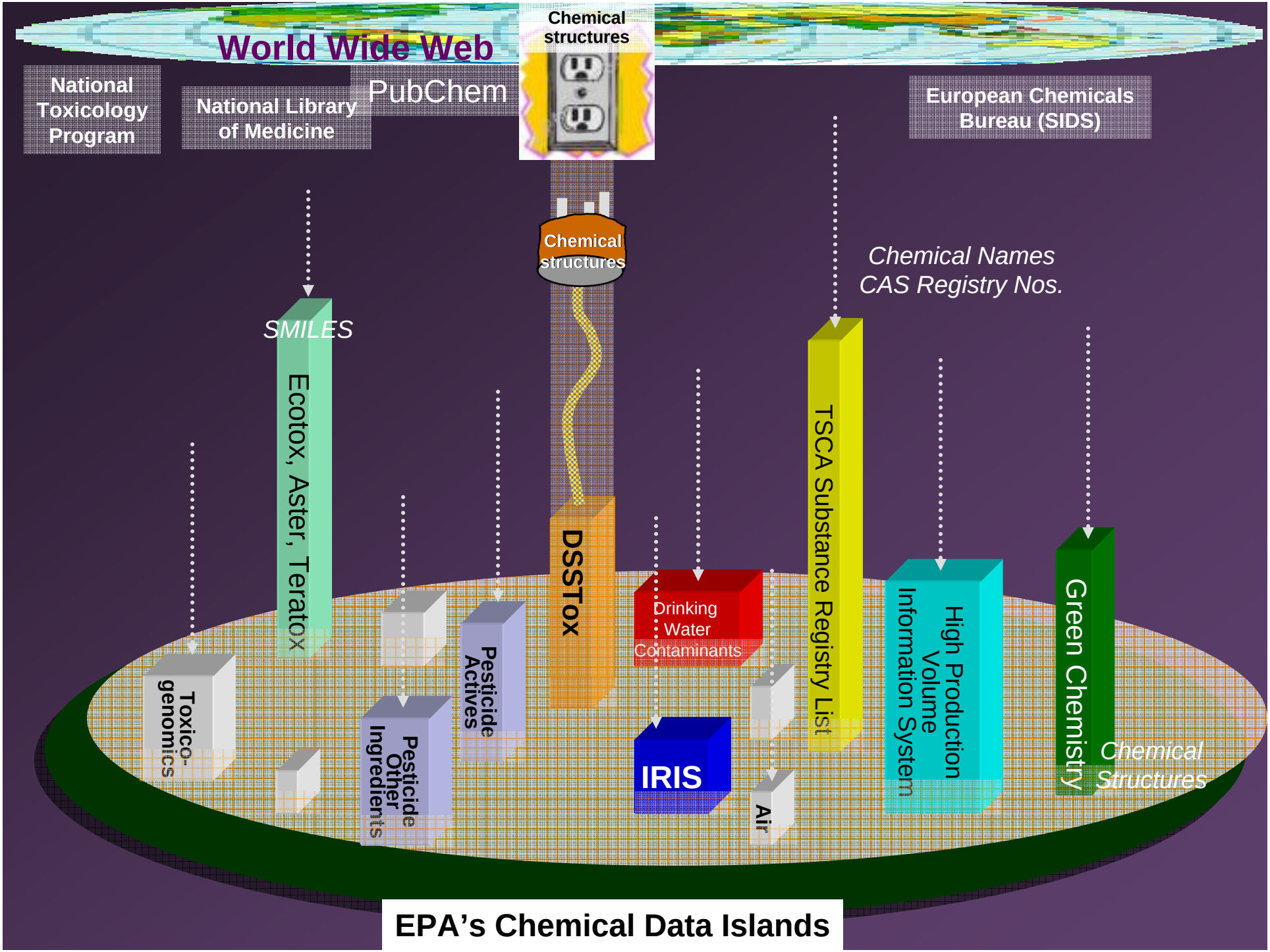
DSSTox Standard Chemical Fields:



DSSTox Master Structure-Index File

8804 total records

	#	NAMEID	DSSTox Master_v1a	
Published Databases	1	CPDBAS	v3b_1481	Carcinogenic Potency Database
	2	DBPCAN	v3b_209	EPA Disinfection By-products Cancer Predictions
	3	EPAFHM	v3b_617	EPA Fathead Minnow Acute toxicity
	4	NCTRER	v3b_232	NCTR Estrogenic Activity
	5	FDAMDD	v2b_1217	FDA Drug Maximum Daily Dose
Published Structure-Index Files	6	HPVCSI	v1a_3548	EPA High Production Volume SI
	7	NTPBSI	v1a_2415	NTP Bioassay SI
	8	IRISSI	v1a_544	EPA IRIS SI
	9	NTPHTS	v1a_1408	NTP High Throughput Screening SI
Databases in development	10	IMMTOX	v1a_87	*NTP Immunotoxicity Testing Battery*
	11	ECODEM	v1a_399	*DEMETRA Pesticides Ecotoxicity*
	12	NTPGTZ	v1a_1931	*NTP Genetic Toxicity - Zeiger*
	13	NCTRAR	v1a_202	*NCTR Androgenic Activity*
Unpublished Structure-Index Files	14	CEBSSI	v1a_20	*NIEHS/NTP CEBS SI*
	15	ICCVAM	v1a_87	*ICCVAM Endocrine Disruption Set SI*
	16	EPAPTC	v1a_826	*EPA Pesticide ToxCast Candidates SI*
	17	EPADNT	v1a_82	*EPA Developmental Neurotox SI*
	18	EPAPAI	v1a_873	*EPA Pesticide Active Ingredients SI*
	19	EPADWC	v1a_66	*EPA Drinking Water Contaminants SI*
	20	EPAPOI	v1a_441	*EPA Pesticide Other Ingredients SI*



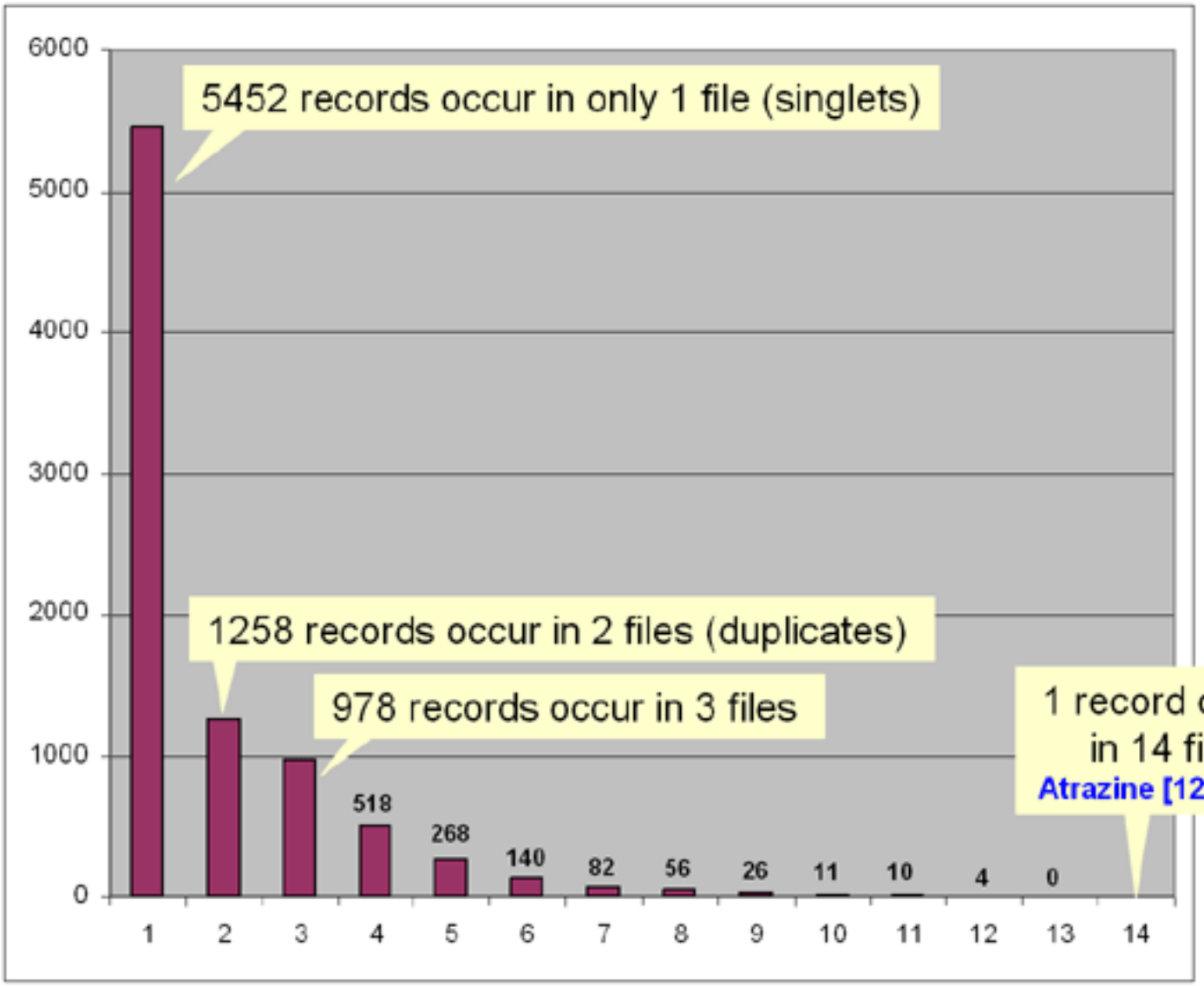
EPA's Chemical Data Islands

DSSTox Master Structure-Index File:

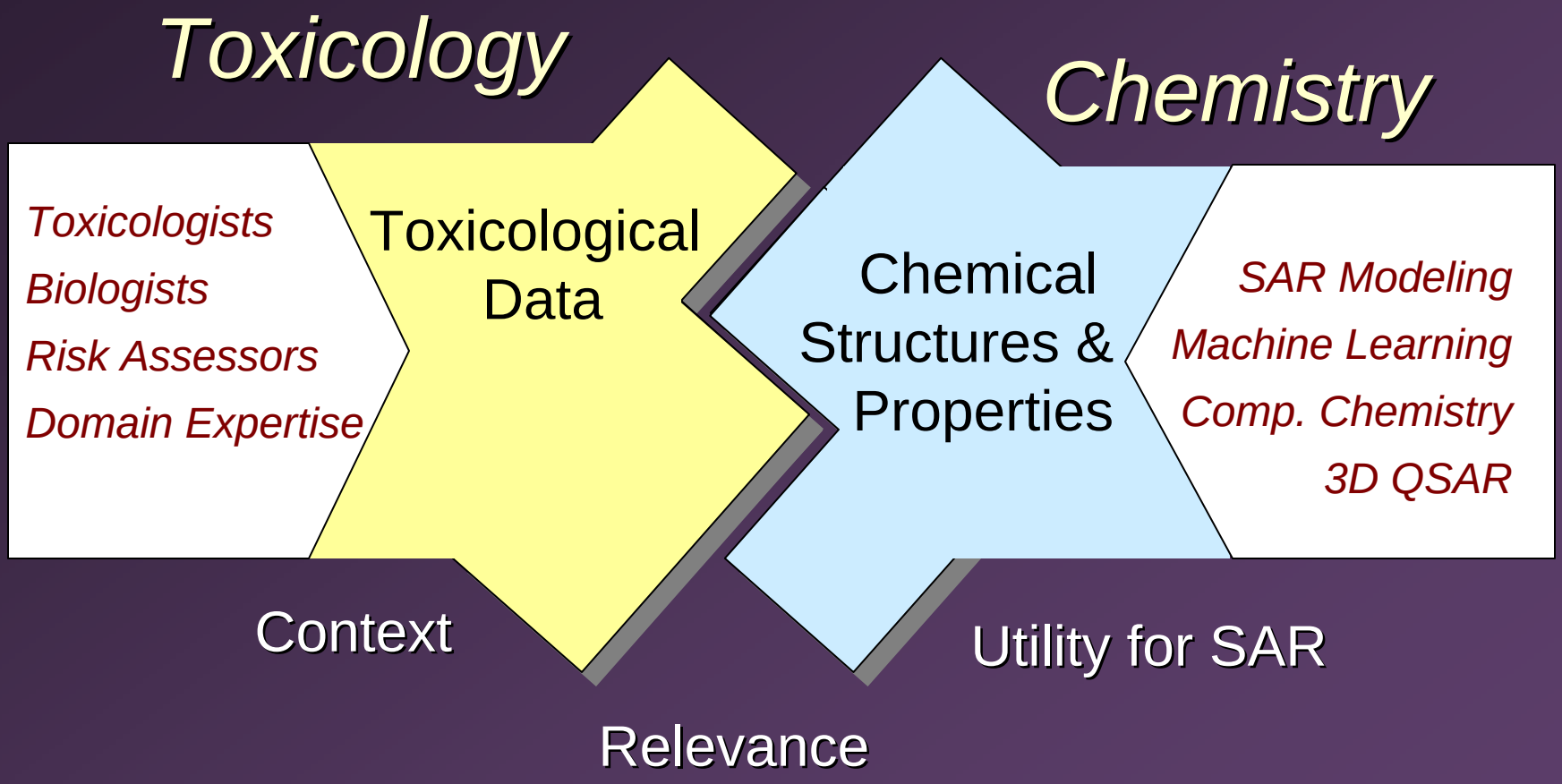
Purpose:
to consolidate, manage, and ensure quality and uniformity of the chemical and substance information spanning all DSSTox Databases and Structure-Index files

I. DSSTox Standard Chemical Fields			*II. DSSTox NAMEID File Incidence Columns (DSSTox_ID)*					*I. DSSTox Standard Chemical Fields*		
STRUCTURE	**DSSTox _CID**	**DSSTox _SID**	**CPDBAS _v3b_1481**	**DBPCAN _v3b_219**	**EPAFHM _v3b_617**	**…**	**Total_File Incidence**	**STRUCTURE _Formula**	**STRUCTURE_ Molecular Weight**	**cont.**
Structure 1	1	1	1	blank	blank	…	1	C11H9N3	183.2122	…
Structure 2	2	2	2		56	…	7	C2H4O	44.0526	…
…	…	…	…	…	…	…	…	…	…	…
Structure 662	614	1085	1085		345	…	3	C24H32O4	384.5085	
…	…	…	…	…	…	…	…	…	…	…
Structure 1625	1513	1513		32	291	…	6	C4H8O	72.1057	
…	…	…	…	…	…	…	…	…	…	…

Total # Chemical Records
DSSToxMaster_v1a

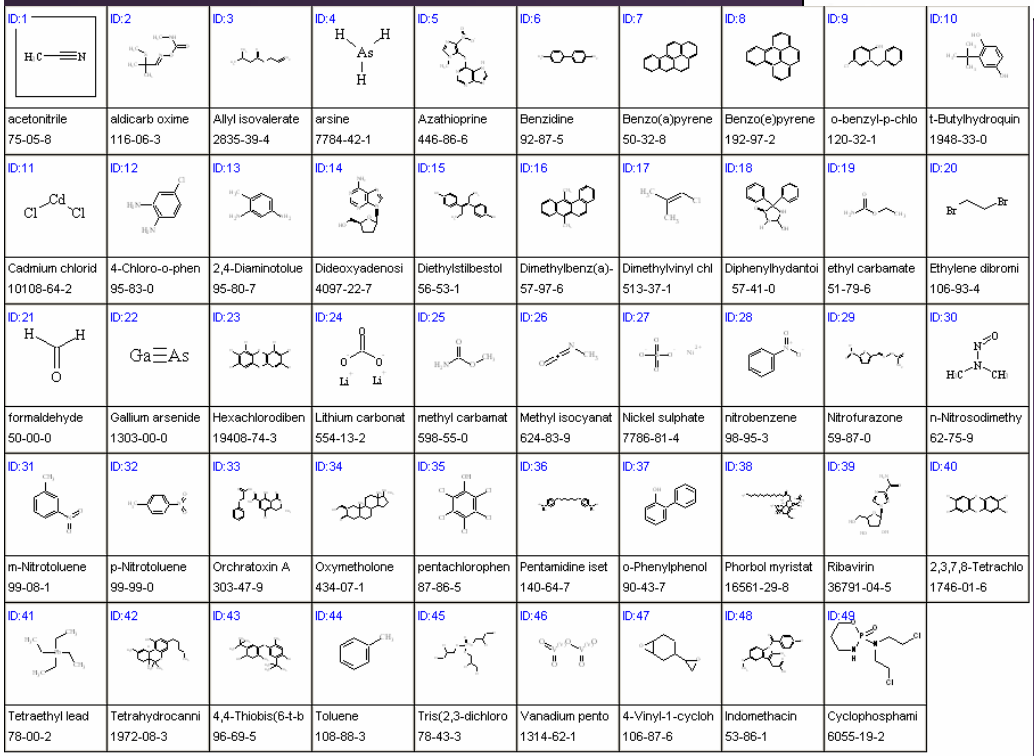


DSSTox Database Design:



IMMTOX: Immunotoxicity Test Battery

- 88 chemicals
- 18 immunotox measures
- Summary calls
- Usage categories



ID:1	ID:2	ID:3	ID:4	ID:5	ID:6	ID:7	ID:8	ID:9	ID:10
acetonitrile 75-05-8	aldicarb oxime 116-06-3	Allyl isovalerate 2835-39-4	arsine 7784-42-1	Azathioprine 446-86-6	Benzidine 92-87-5	Benzo(a)pyrene 50-32-8	Benzo(e)pyrene 192-97-2	o-benzyl-p-chlo 120-32-1	t-Butylhydroquin 1948-33-0
ID:11	ID:12	ID:13	ID:14	ID:15	ID:16	ID:17	ID:18	ID:19	ID:20
Cadmium chlorid 10108-64-2	4-Chloro-o-phen 95-83-0	2,4-Diaminotolue 95-80-7	Dideoxyadenosi 4097-22-7	Diethylstilbestol 56-53-1	Dimethylbenz(a)- 57-97-6	Dimethylvinyl chl 513-37-1	Diphenylhydantoi 57-41-0	ethyl carbamate 51-79-6	Ethylene dibromi 106-93-4
ID:21	ID:22	ID:23	ID:24	ID:25	ID:26	ID:27	ID:28	ID:29	ID:30
formaldehyde 50-00-0	Gallium arsenide 1303-00-0	Hexachlorodiben 19408-74-3	Lithium carbonat 554-13-2	methyl carbamat 598-55-0	Methyl isocyanat 624-83-9	Nickel sulphate 7786-81-4	nitrobenzene 98-95-3	Nitrofurazone 59-87-0	n-Nitrosodimethy 62-75-9
ID:31	ID:32	ID:33	ID:34	ID:35	ID:36	ID:37	ID:38	ID:39	ID:40
m-Nitrotoluene 99-08-1	p-Nitrotoluene 99-99-0	Orchratoxin A 303-47-9	Oxymetholone 434-07-1	pentachlorophen 87-86-5	Pentamidine iset 140-64-7	o-Phenylphenol 90-43-7	Phorbol myristat 16561-29-8	Ribavirin 36791-04-5	2,3,7,8-Tetrachlo 1746-01-6
ID:41	ID:42	ID:43	ID:44	ID:45	ID:46	ID:47	ID:48	ID:49	
Tetraethyl lead 78-00-2	Tetrahydrocanni 1972-08-3	4,4-Thiobis(6-t-b 96-69-5	Toluene 108-88-3	Tris(2,3-dichloro 78-43-3	Vanadium pento 1314-62-1	4-Vinyl-1-cycloh 106-87-6	Indomethacin 53-86-1	Cyclophosphami 6055-19-2	

FUNDAMENTAL AND APPLIED TOXICOLOGY **10,** 2–19 (1988)

METHODS EVALUATION

Development of a Testing Battery to Assess Chemical-Induced Immunotoxicity: National Toxicology Program's Guidelines for Immunotoxicity Evaluation in Mice

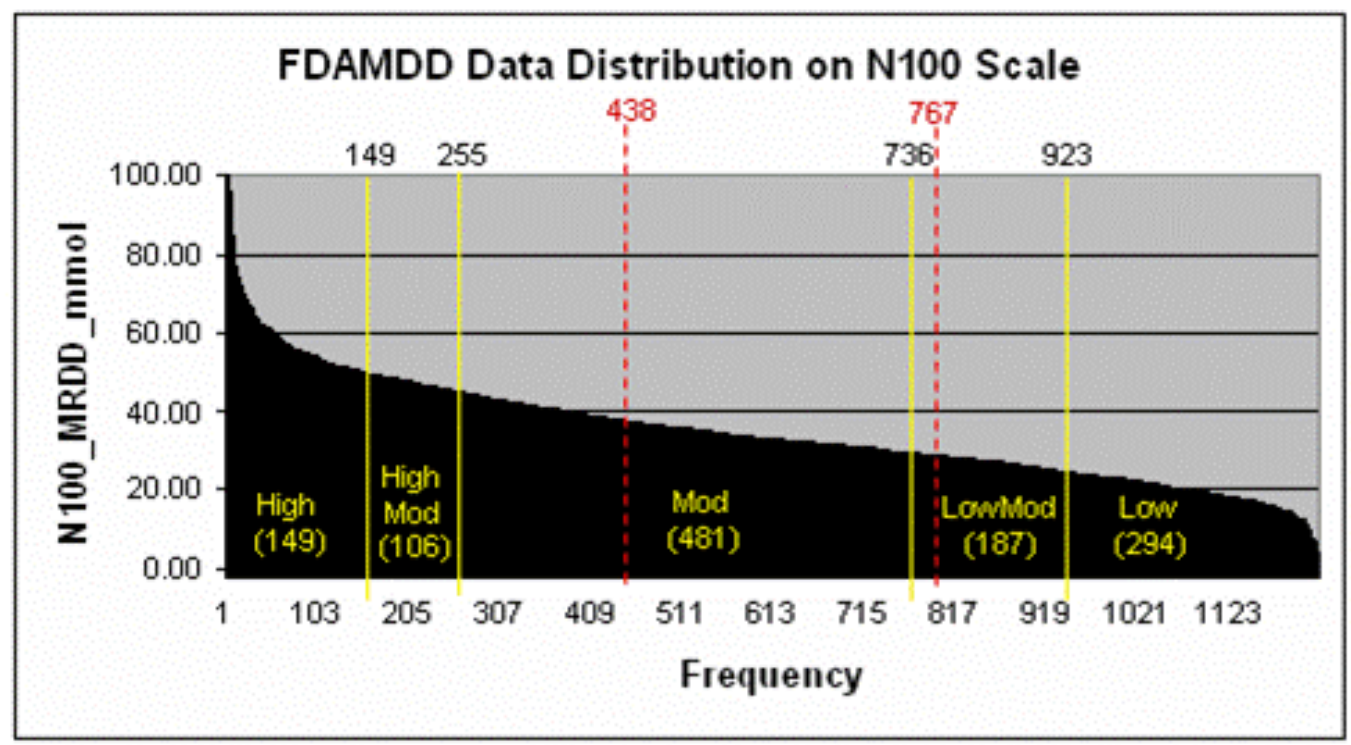
MICHAEL I. LUSTER,* ALBERT E. MUNSON,† PETER T. THOMAS,‡
MICHAEL P. HOLSAPPLE,† JAMES D. FENTERS,‡ KIMBER L. WHITE, JR.,†
LLOYD D. LAUER,§ DORI R. GERMOLEC,* GARY J. ROSENTHAL,* AND JACK H. DEAN§

*Systemic Toxicology Br
P.O. Box 12233, Resear
Medical College of Vi
Division, IIT Research I
Toxicology, Departr

- AntibodyResponse
- NaturalKillerCells
- LymphocyteProliferation
- MixedLeukocyteResponse
- LeukocyteCount
- ThymusBodyWt
- SpleenBodyWt
- Lipopolysaccaride
- DelayedTypeHypersensitivity
- CytotoxicTLymphocyte
- SurfaceMarkers
- ContactSensitization
- HostResistanceAssays (6)

FDAMDD: FDA's Center for Drug Evaluation & Research - Maximum (Recommended) Daily Dose Database

MRDD values for 1217 pharmaceuticals in mg/kg-body weight (bw)/day extracted from public literature sources (D. Benz, E. Matthews, N. Kruhlac, J. Contera)



N100 Scale:
High (100-50)
HighMod (50-45)
Mod (45-30)
LowMod (30-25)
Low (25-0)

MCASE (approx):
High (100-37.8)
Mod (37.8-29.3)
Low (29.3-0)

CPDBAS_v3b_1481

STRUCTURE
DSSTox_CID
DSSTox_SID
DSSTox_ID_FileName
STRUCTURE_Formula
STRUCTURE_MolecularWe
STRUCTURE_ChemicalTyp
STRUCTURE_TestedForm
_DefinedOrganic
STRUCTURE_Shown
TestSubstance_ChemicalNa
TestSubstance_CASRN
TestSubstance_Description
ChemicalNote
STRUCTURE_ChemicalNan
_IUPAC
STRUCTURE_SMILES
STRUCTURE_Parent_SMIL
STRUCTURE_InChI
StudyType
Endpoint
Species

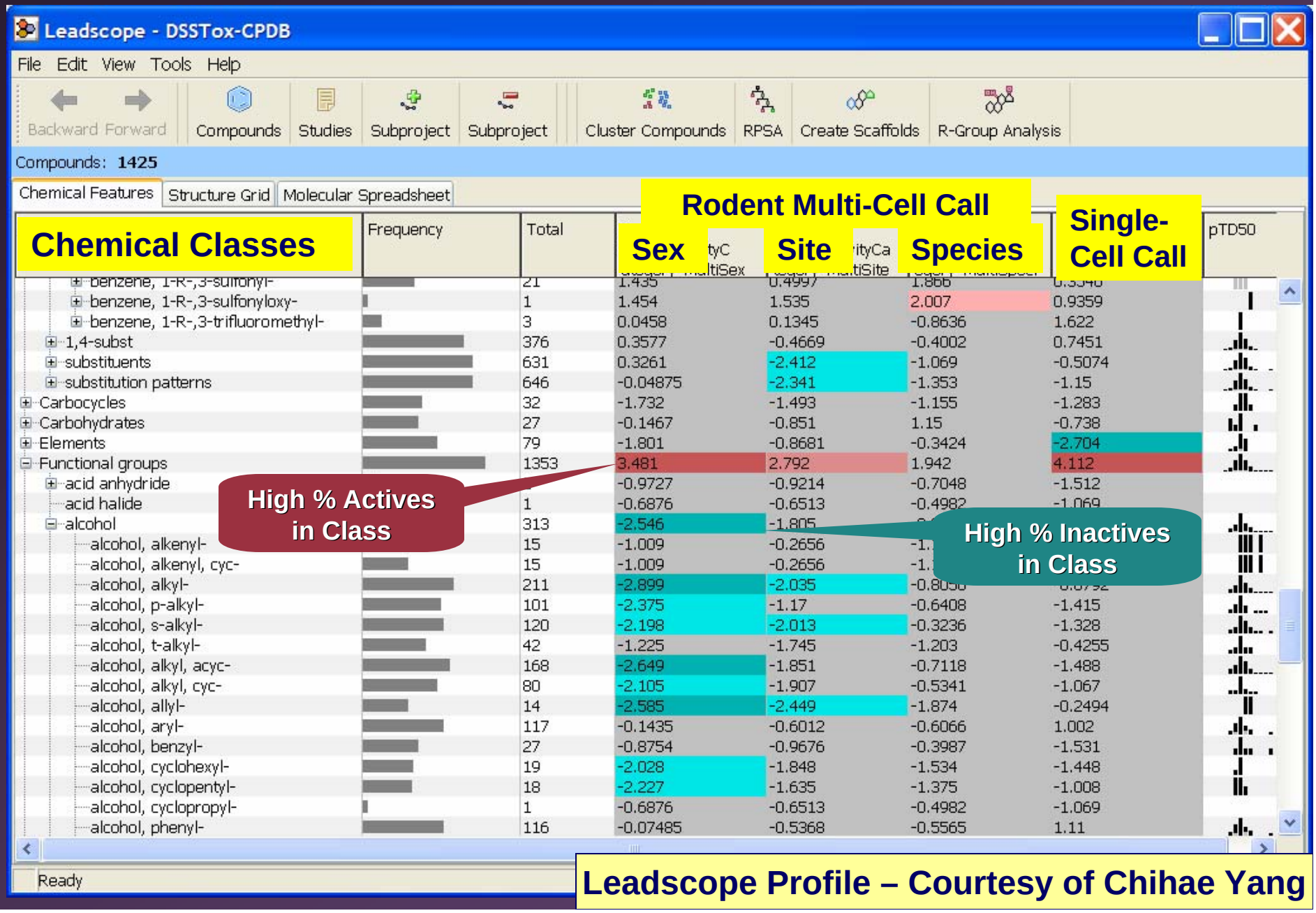
SAL_CPDB
TD50_Rat_mg/kg/day
TD50_Rat_mmol/kg/day
TargetSites_Rat_Male, Fema oth Sex
TD50_Mouse_mg/kg/day
TD50_Mouse_mmol/kg/day
TargetSites_Mouse_Male, Female, Both
TD50_Hamster_mg/kg/day
TD50_Hamster_mmol/kg/day
TargetSites_Hamster_Male, Female, Both
TD50_Dog_mg/kg/day
TargetSites_Dog
TD50_Rhesus_mg/kg/day
TargetSites_Rhesus
TD50_Cynomolgus_mg/kg/day
TargetSites_Cynomolgus
ActivityCategory_SingleCellCall
ActivityCategory_MultiCellCall
ToxicityNote
NTP_TechnicalReport
Website_URL

0
1

multisite
multisex
multispecies

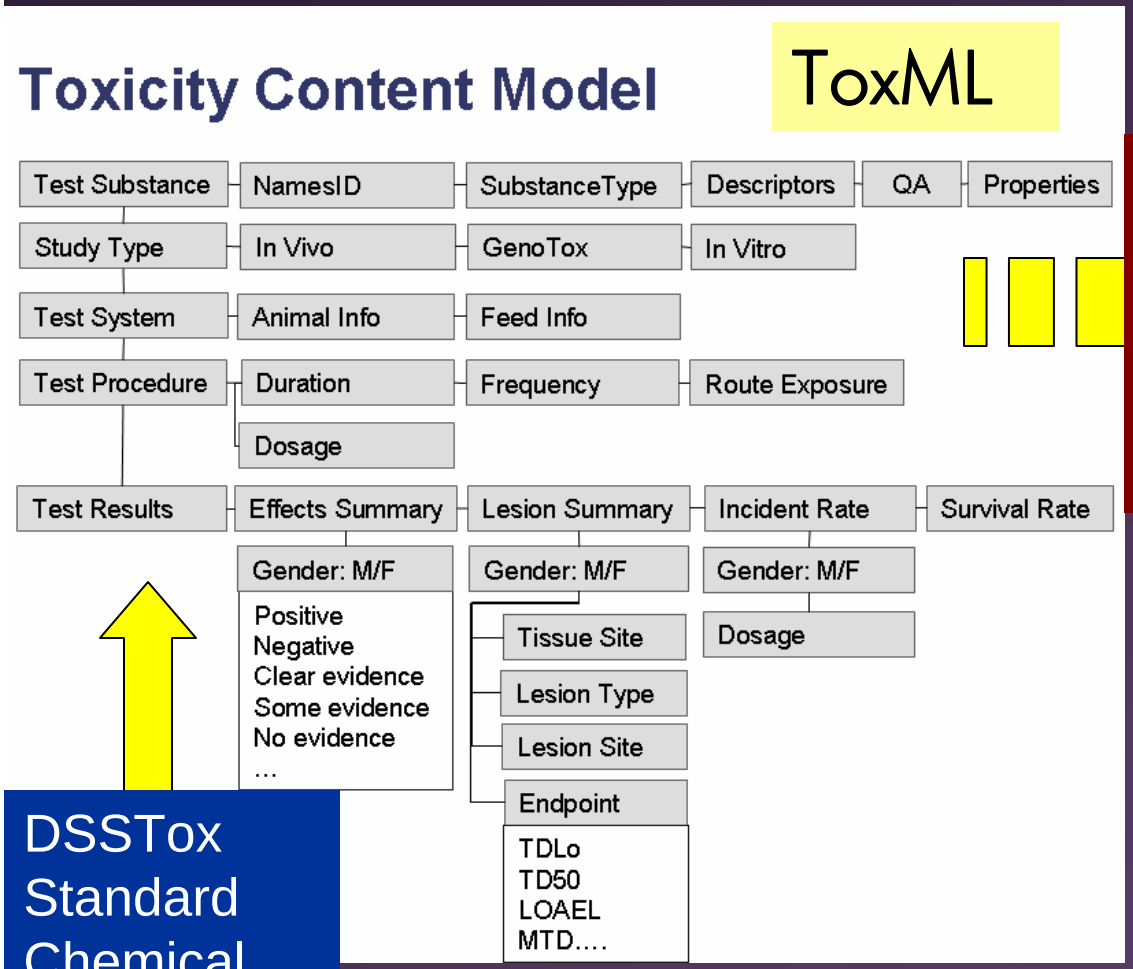
adr = adrenal gland;
bon = bone;
cli = clitoral gland;
eso = esophagus;
ezy = ear/Zymbal's gland;
gal = gall bladder;
hag = harderian gland;
hmo = hematopoietic system;
kid = kidney;
lgi = large intestine;
liv = liver;
lun = lung;
meo = mesovarium;
mgl = mammary gland;
mix = mixture;
myc = myocardium;
nas = nasal cavity
nrv = nervous system;
orc = oral cavity
ova = ovary;
pan = pancreas;
per = peritoneal cavity;
pit = pituitary gland;
pre = preputial gland;
pro = prostate;
ski = skin;
smi = small intestine;
spl = spleen;
sto = stomach;
sub = subcutaneous tissue;
ng animals;
vag = vagina;
vsc = vascular system.

Chemical-Biological Profiling of CPDBAS Activities



Leadscope Profile – Courtesy of Chihae Yang

Toxicity Experimental Data → Summary Data:



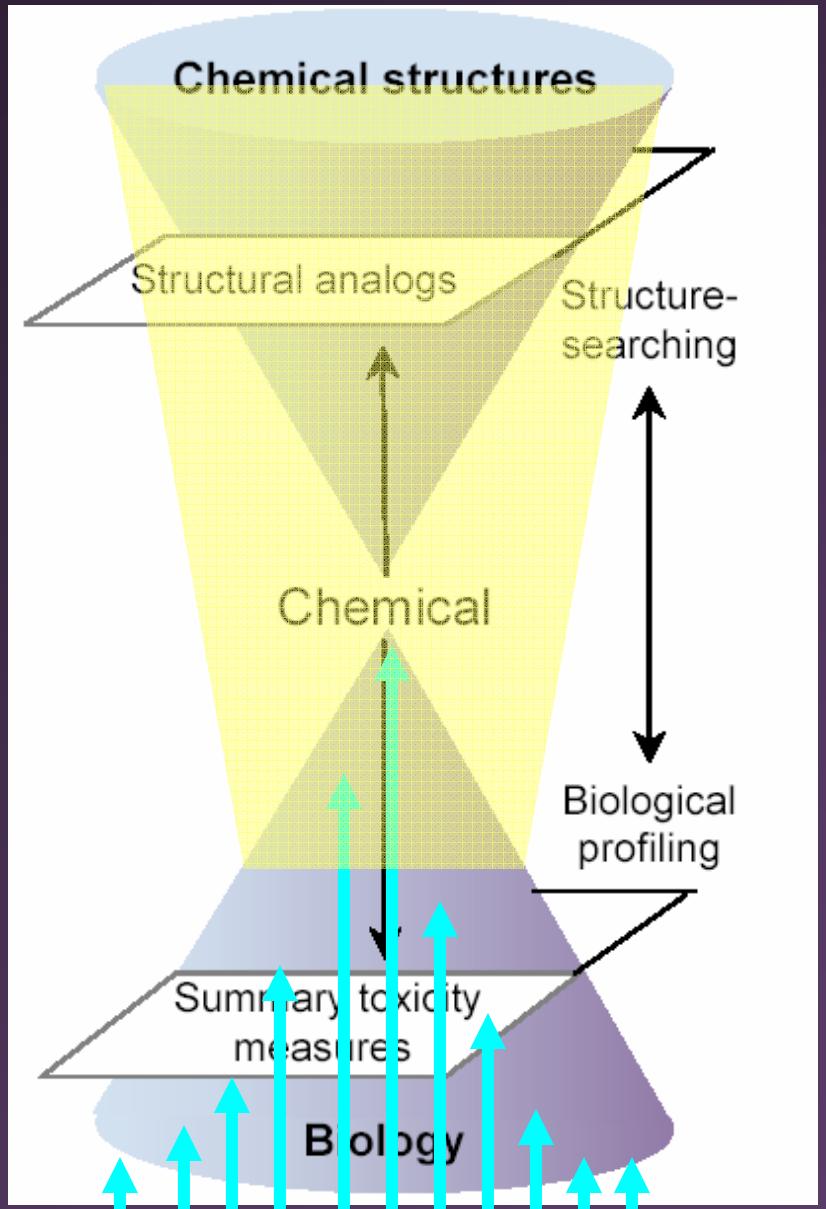
Intermediate toxicity classifications for SAR

+

−

- Activity categories
- Potency categories
- Mode of action categories
- Summary calls

DSSTox Summary Toxicity Data Files



ToxML

Leadscope Data Mining

Leadscope QSAR Databases

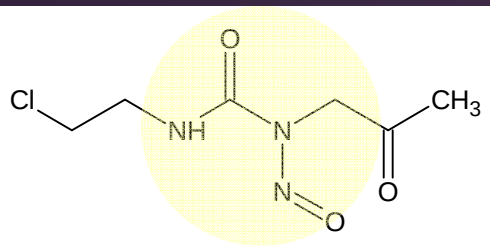
Chemical Errors in Toxicity Information:

- Name is misspelled
- CAS is invalid or retired
- CAS and name do not agree
- Name and structure do not agree
- Name is insufficient for structure assignment
- Insufficient description of substance

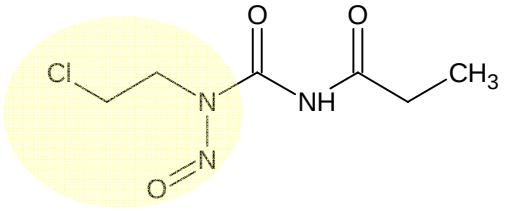
Carcinogenic Potency Database:
Hamster Carcinogenicity Results

From CPDB Hamster Table:

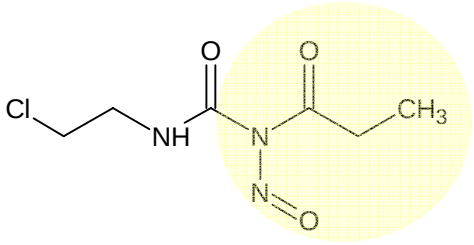
N-Nitroso-oxopropylchloroethylurea NOCAS



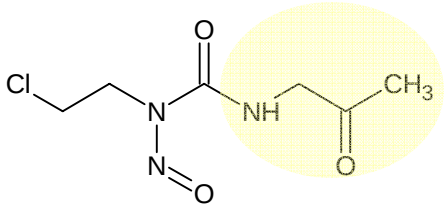
3-(2-chloroethyl)-1-nitroso-1-(2-oxopropyl)urea



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



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



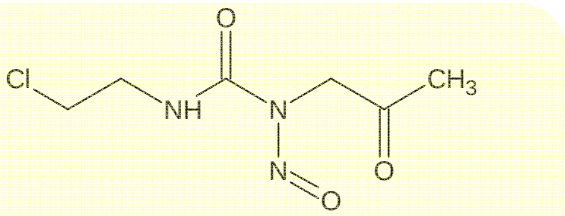
1-(2-chloroethyl)-1-nitroso-3-(2-oxopropyl)urea

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

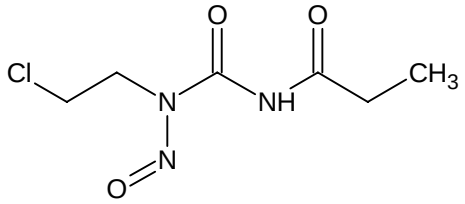
Carcinogenic Potency Database:
Hamster Carcinogenicity Results

From CPDB Hamster Table:

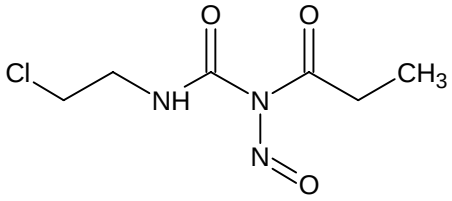
N-Nitroso-oxopropylchloroethylurea NOCAS



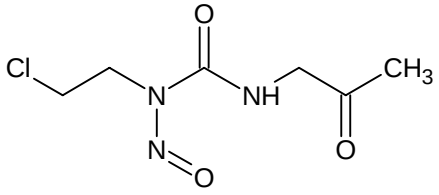
3-(2-chloroethyl)-1-nitroso-1-(2-oxopropyl)urea



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



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



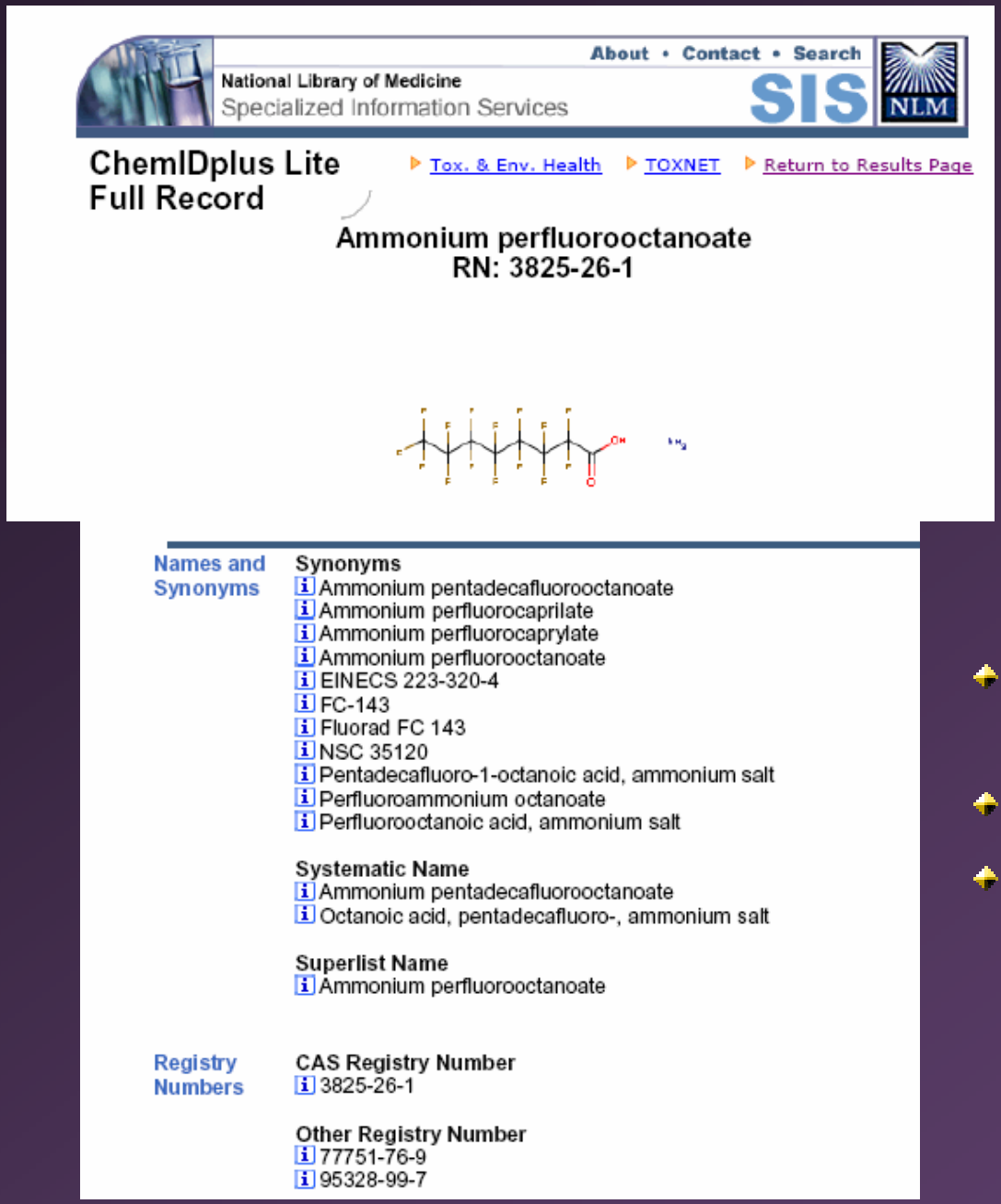
1-(2-chloroethyl)-1-nitroso-3-(2-oxopropyl)urea

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

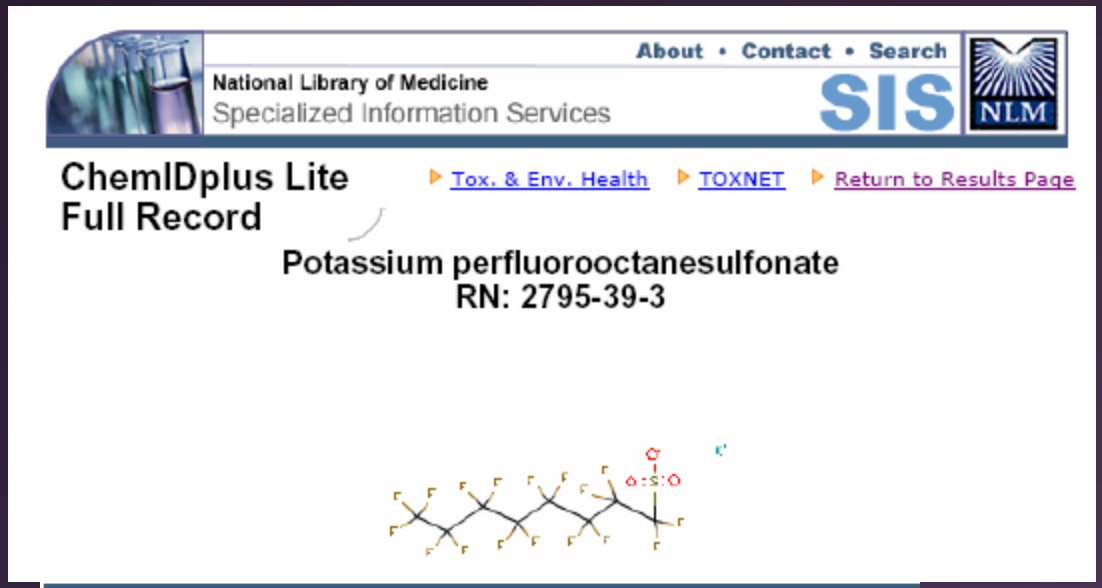
Perfluoroalkylacids (PFAAs):

- 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 concern to EPA
- PFOA and PFOS have undergone extensive toxicity testing
  - *Hepatotoxic*
  - *Developmental toxicants*
  - *Immunotoxic*



PFOA

- Major synthetic pathway is telemerization
- Yields 98% pure linear form
- Verified by primary manufacturer (Dupont) and NMR



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

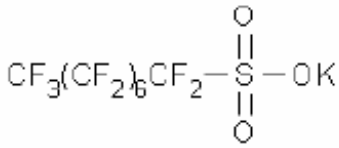
- Major synthetic pathway is electrochemical fluorination
- Yields approx 60-80% linear form, with significant non-linear contamination
- Verified by primary manufacturer (3M) and NMR

©www.sigma-aldrich.com

Product Name or No.

77282 Heptadecafluorooctanesulfonic acid potassium salt

Fluka purum, ≥98.0% (T)



Synonym	Perfluorooctanesulfonic acid potassium salt Potassium heptadecafluoro-1-octanesulfonate
Molecular Formula	$CF_3(CF_2)_7SO_3K$
Molecular Weight	538.22
CAS Number	2795-39-3
Beilstein Registry Number	3864579
EG/EC Number	2205287
MDL number	MFCD00066407

Expand/Collapse All

Price and Availability

Product Number	Your Price USD	Available to Ship		Quantity	Actions
77282-10G	62.70	02/22/2007	details...		
77282-50G	238.00	02/22/2007	details...		

Properties

grade	purum
assay	≥98.0% (T)
mp	277-280 °C(lit.)

Safety

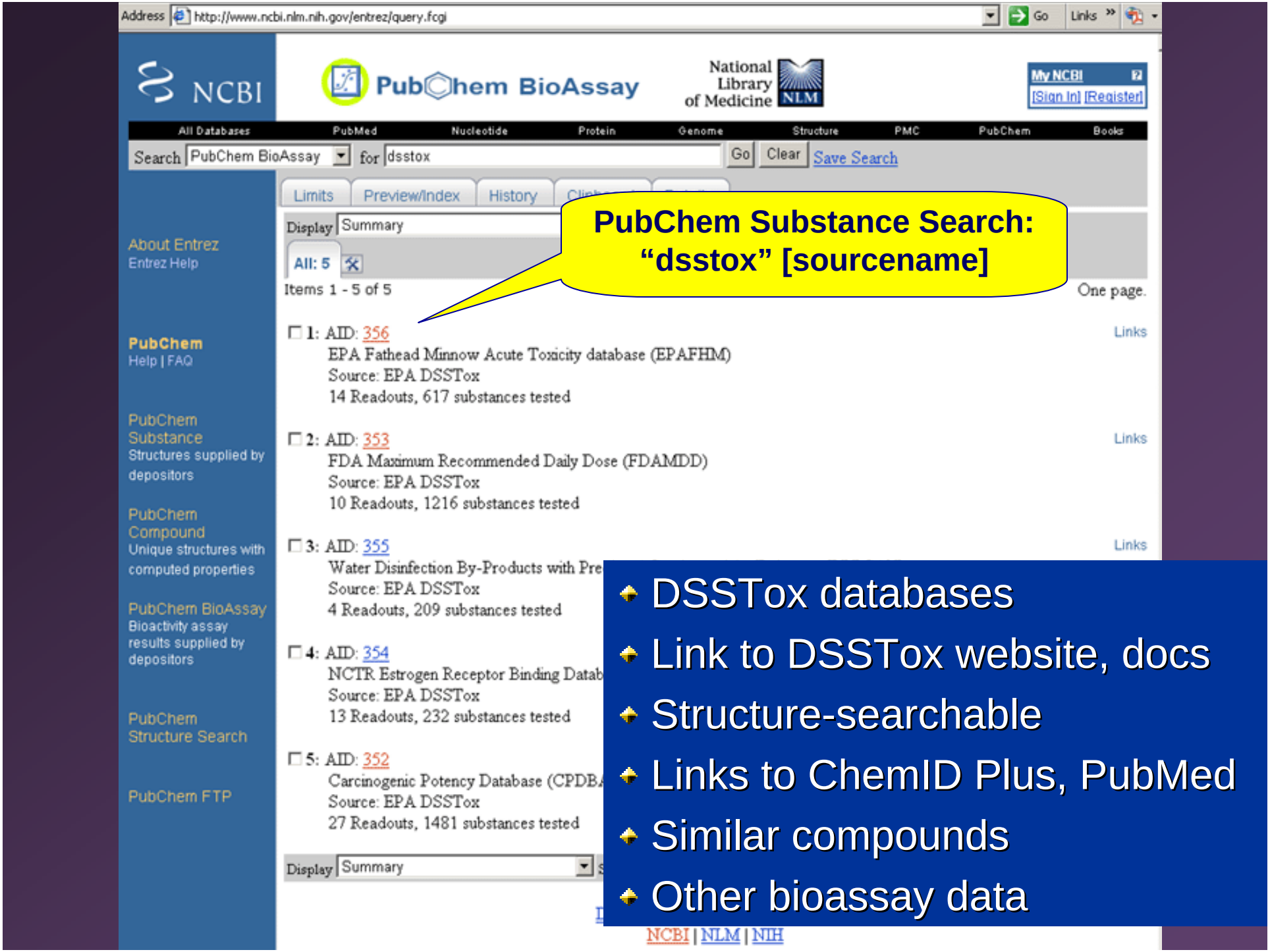
Hazard Codes	Xn
Risk Statements	22-36/37/38
Safety Statements	26
WGK Germany	3

Related Categories

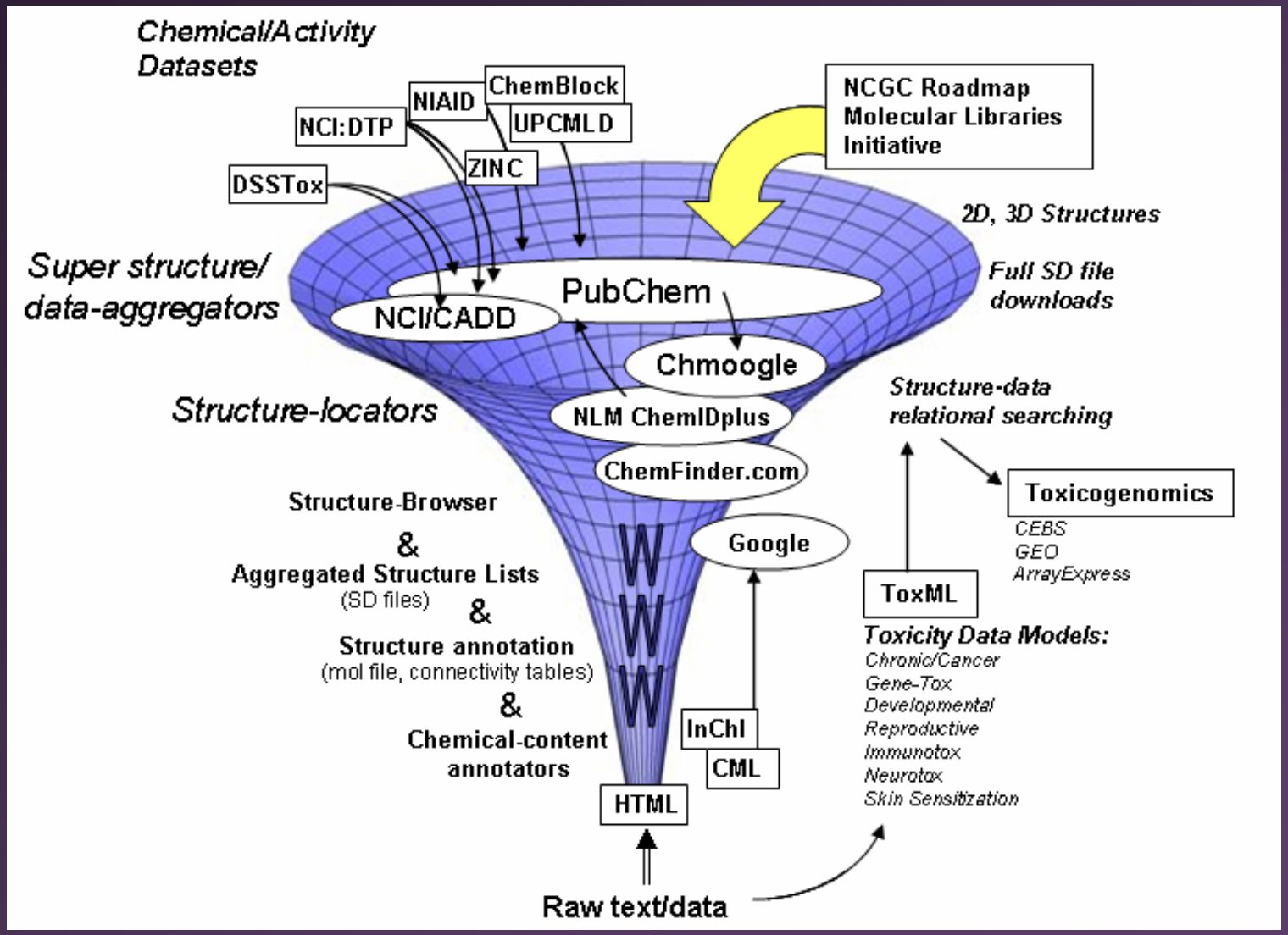
... Sulfur Compounds > Sulfonic/Sulfinic Acids

PFOS

- Fluka is secondary chemical distributor
- Listed as 98.0% pure (T)
- (T) indicates titration method which confirms only empirical formula
- Listed as linear form by both structure and CASRN

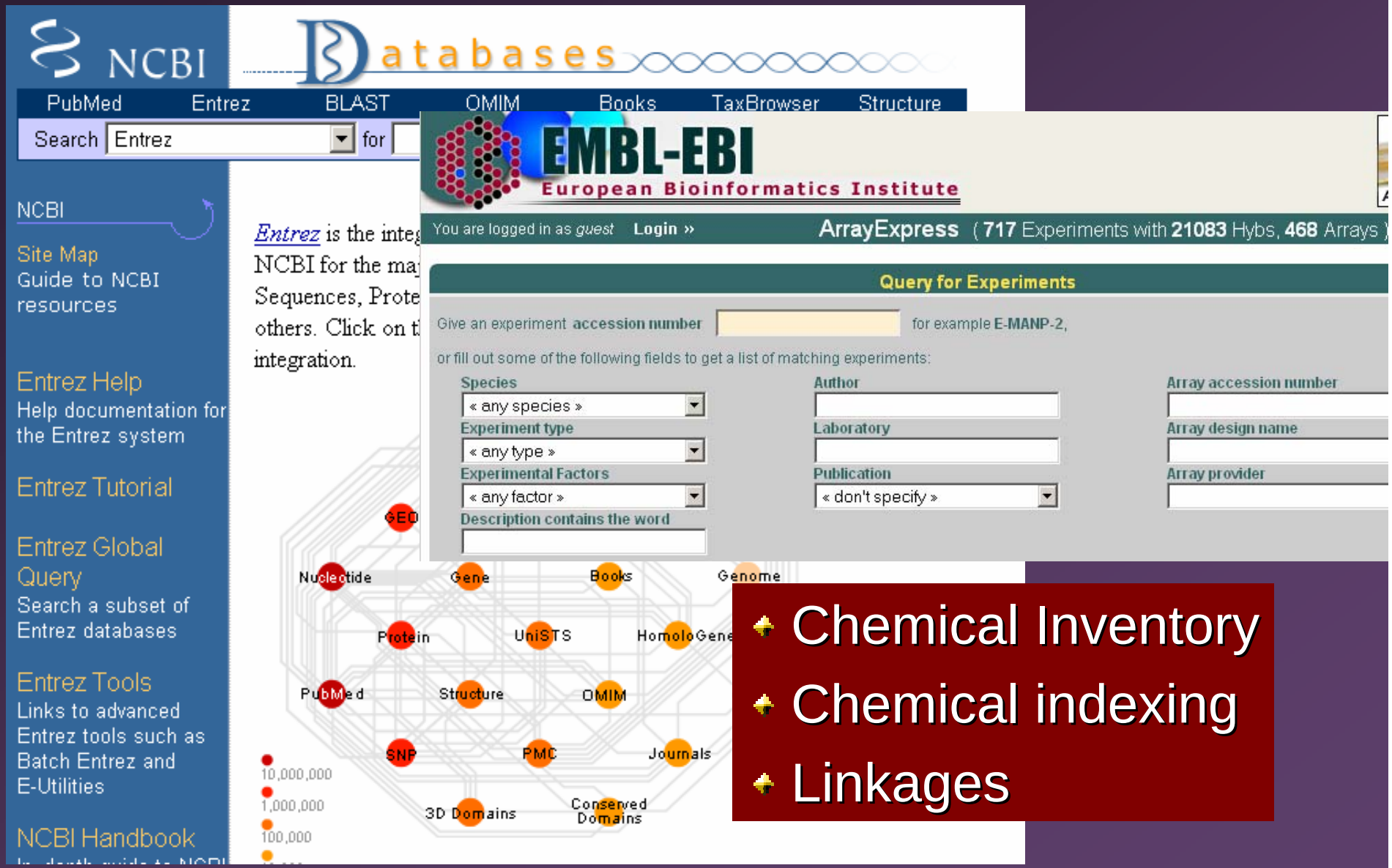

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi
Go
Links
NCBI
PubChem BioAssay
National Library of Medicine
NLM
My NCBI
[Sign In] [Register]
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
PMC
PubChem
Books
Search PubChem BioAssay for dsstox
Go
Clear
Save Search
Limits
Preview/Index
History
Display Summary
All: 5
Items 1 - 5 of 5
One page.
About Entrez
Entrez Help
PubChem
Help | FAQ
PubChem Substance
Structures supplied by depositors
PubChem Compound
Unique structures with computed properties
PubChem BioAssay
Bioactivity assay results supplied by depositors
PubChem Structure Search
PubChem FTP
1: AID: 356
EPA Fathead Minnow Acute Toxicity database (EPAFHM)
Source: EPA DSSTox
14 Readouts, 617 substances tested
Links
2: AID: 353
FDA Maximum Recommended Daily Dose (FDAMDD)
Source: EPA DSSTox
10 Readouts, 1216 substances tested
Links
3: AID: 355
Water Disinfection By-Products with Pre
Source: EPA DSSTox
4 Readouts, 209 substances tested
Links
4: AID: 354
NCTR Estrogen Receptor Binding Datab
Source: EPA DSSTox
13 Readouts, 232 substances tested
5: AID: 352
Carcinogenic Potency Database (CPDBA
Source: EPA DSSTox
27 Readouts, 1481 substances tested
Display Summary
NCBI | NLM | NIH
PubChem Substance Search: "dsstox" [sourcename]
DSSTox databases
Link to DSSTox website, docs
Structure-searchable
Links to ChemID Plus, PubMed
Similar compounds
Other bioassay data

Chemical Structure Searching of Biological Information on the Internet

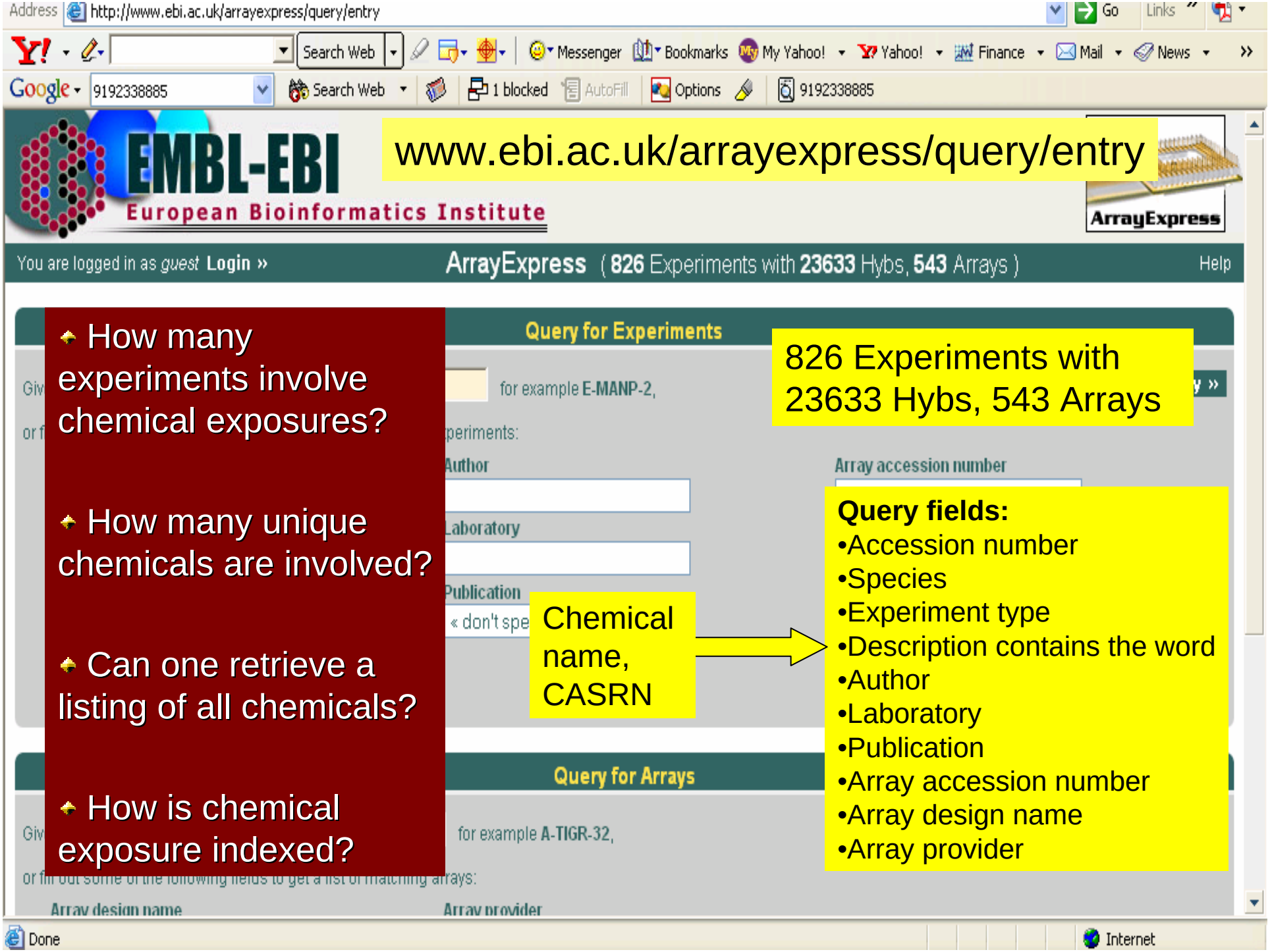


Richard A.M., Gold, L.S., Nicklaus, M.C. (2006) Chemical structure indexing of toxicity data on the internet: Moving towards a flat world. Curr. Opinion Drug Disc. Develop., 9(3): 314-325.

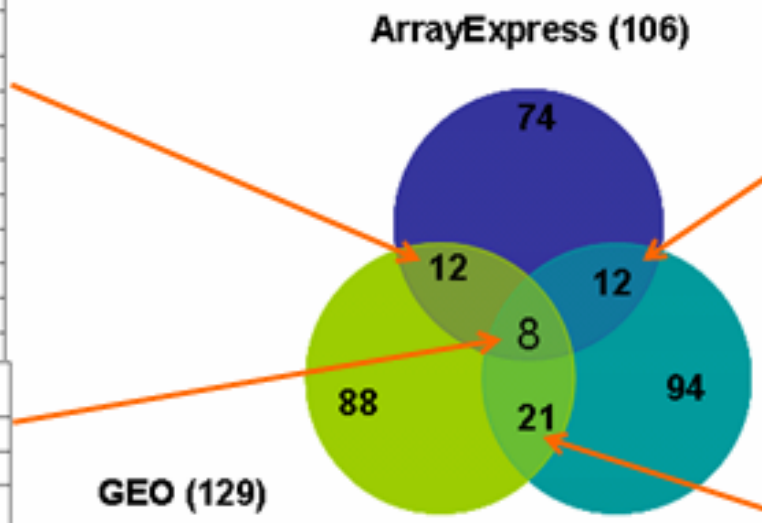
Public Genomic Databases



- Chemical Inventory
- Chemical indexing
- Linkages


Address http://www.ebi.ac.uk/arrayexpress/query/entry
Go
Links
Search Web
Messenger
Bookmarks
My Yahoo!
Yahoo!
Finance
Mail
News
Google
9192338885
Search Web
1 blocked
AutoFill
Options
9192338885
EMBL-EBI
European Bioinformatics Institute
www.ebi.ac.uk/arrayexpress/query/entry
ArrayExpress
You are logged in as guest Login »
ArrayExpress (826 Experiments with 23633 Hybs, 543 Arrays)
Help
How many experiments involve chemical exposures?
How many unique chemicals are involved?
Can one retrieve a listing of all chemicals?
How is chemical exposure indexed?
Query for Experiments
for example E-MANP-2,
Author
Laboratory
Publication
« don't spe
Array accession number
826 Experiments with 23633 Hybs, 543 Arrays
Chemical name, CASRN
Query fields:
•Accession number
•Species
•Experiment type
•Description contains the word
•Author
•Laboratory
•Publication
•Array accession number
•Array design name
•Array provider
Query for Arrays
for example A-TIGR-32,
Array design name
Array provider
Done
Internet

Overlap between Genomic Repository Chemical Space



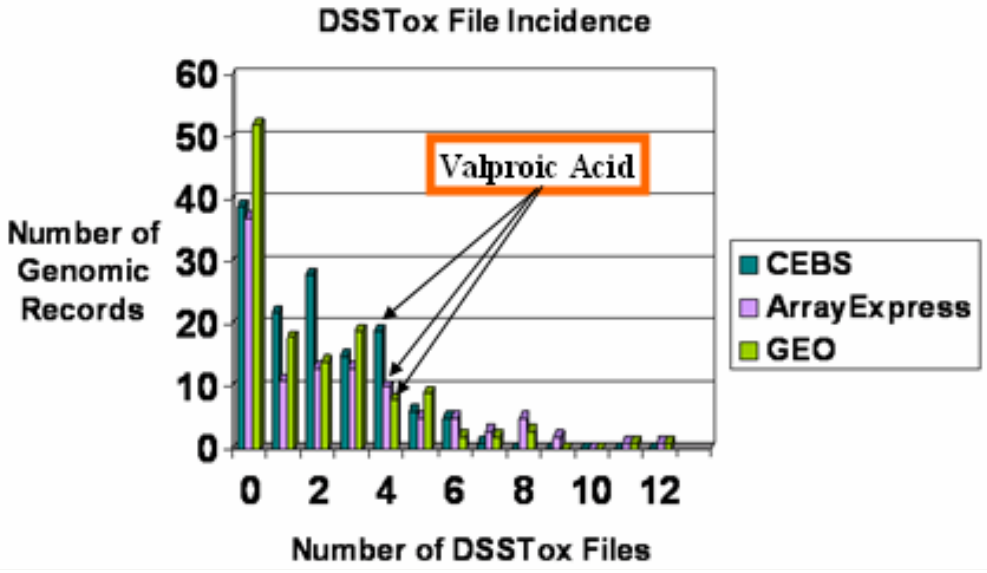
Chemical Overlap Between ArrayExpress and GEO	
17beta-Estradiol	50-28-2
Genistein	446-72-0
Dexamethazone	50-02-2
Cycloheximide	66-81-9
Retinoic Acid	302-79-4
Hydrogen peroxide	7722-84-1
Terephthalic acid,TPA	100-21-0
Fenpropimorph	67306-03-0
5-Fluorouracil	51-21-8
Liothyronine;	6893-02-3
Mitomycin C	50-07-7
Paclitaxel	33069-62-4

Chemical Overlap between ArrayExpress, GEO, and CEBS	
cis-Dichlorodiamine platinum	15663-27-1
Clofibrate	637-07-0
Clotrimazole	23593-75-1
Ethanol	64-17-5
Ethinyl estradiol	57-63-6
Lipopolysaccharide (LPS)	NOCAS
Miconazole	22916-47-8
Valproic acid	99-66-1

Chemical Overlap between ArrayExpress and CEBS	
2-Acetylaminofluorene	53-96-3
4-Acetylaminofluorene	28322-02-3
Acetaminophen (4-hydroxyacetanilide)	103-90-2
Benitradine	88133-11-3
Butylated hydroxytoluence (BHT)	128-37-0
Doxylamine succinate	562-10-7
Etoposide	33419-42-0
Methapyrilene	91-80-5
Methyl viologen	1910-42-5
N-Nitrosodimethylamine	62-75-9
phenobarbital	50-06-6
Tamoxifen	10540-29-1

Chemical Overlap between GEO and CEBS	
1,3-Bis(chloroethyl)-1-nitrosourea	154-93-8
Adriamycin, hydrochloride	25316-40-9
Allyl alcohol	107-18-6
alpha-Naphthyl isothiocyanate	551-06-4
Atorvastatin	134523-00-5
Bezafibrate	41859-67-0
Carbon tetrachloride	56-23-5
Dimethylformamide	68-12-2
Econazole	27220-47-9
Fenofibrate	49562-28-9
Fluconazole	86386-73-4
Fluvastatin	93957-54-1
Gemfibrozil	25812-30-0
Itraconazole	84625-61-6
Lovastatin	75330-75-5
Methotrexate	59-05-2
N-Nitrosodiethylamine	55-18-5
Rifampicin	13292-46-1
SIMVASTATIN	79902-63-9
Sodium arsenite	7784-46-5
Troglitazone	97322-87-7

Overlap between Genomic Chemical Space and DSSTox Chemical Space



Part IV.

Expanded view of

“Chemical analogs”

Predictive Toxicology

SAR Predictions based solely on chemical structures & properties

Toxicity predictions based on HTS or gene expression profiles,
eg Gene-Logic

Chemical “Probes” of biological systems:

“We find that the connection between structure and biological response is not symmetric, with biological response better at predicting chemical structure than vice versa.”

D. Covell and coworkers
NCI Developmental Therapeutics Program
J Chem Inf Model (2006) 46:430-437

Biological spectra analysis: Linking biological activity profiles to molecular structure

PNAS **January 11, 2005** vol. 102 no. 2 **261–266**

CHEMISTRY PHARMACOLOGY

Anton F. Fliri*, William T. Loging, Peter F. Thadeio, and Robert A. Volkmann*†

Pfizer Global Research and Development, Groton, CT 06340

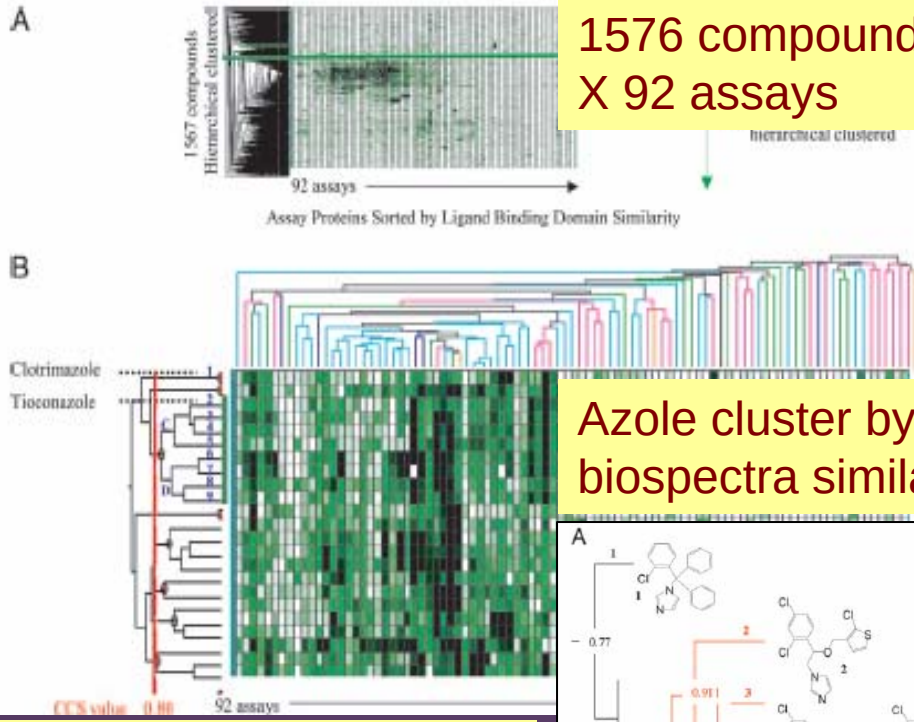
Communicated by Larry E. Overman, University of California, Irvine, CA, October 25, 2004 (received for review September 4, 2004)

Establishing quantitative relationships between molecular structure and broad biological effects has been a longstanding challenge in science. Currently, no method exists for forecasting broad biological activity profiles of medicinal agents even within narrow boundaries of structurally similar molecules. Starting from the premise that biological activity results from the capacity of small organic molecules to modulate the activity of the proteome, we set out to investigate whether descriptor sets could be developed for measuring and quantifying this molecular property. Using a 1,567-compound database, we show that percent inhibition values, determined at single high drug concentration in a battery of *in vitro* assays representing a cross section of the proteome, provide precise molecular property descriptors that identify the structure of molecules. When broad biological activity of molecules is represented in spectra form, organic molecules can be sorted by quantifying differences between biological spectra. Unlike traditional structure–activity relationship methods, sorting of molecules by using biospectra comparisons does not require knowledge of a molecule's putative drug targets. To illustrate this finding, we selected as starting point the biological activity spectra of clotrimazole and tioconazole because their putative target, lanosterol demethylase (CYP51), was not included in the bioassay array. Spectra similarity obtained through profile similarity measurements and hierarchical clustering provided an unbiased means for establishing quantitative relationships between chemical structures and biological activity spectra. This methodology, which we have termed biological spectra analysis, provides the capability not only of sorting molecules on the basis of biospectra similarity but also of predicting simultaneous interactions of new molecules with multiple proteins.

biospectra | proteome | structure–function relationships

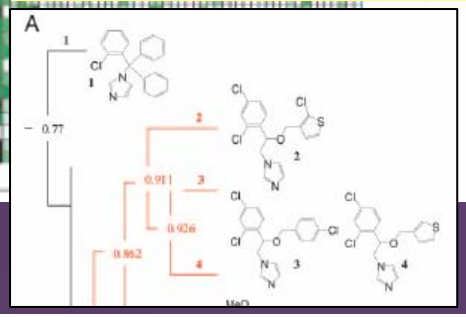
differences in biological environments (8). Considering the complexity of this requirement, computational solutions that precisely link molecular structure to broad biological response are currently not possible (9, 10). We report here an approach to structure–function studies that is based on measurements of the capacity of molecules to interact with the proteome (11).

Translation of Chemical Property Information into Biological Activity Spectra



1576 compounds
X 92 assays

Azole cluster by biospectra similarity

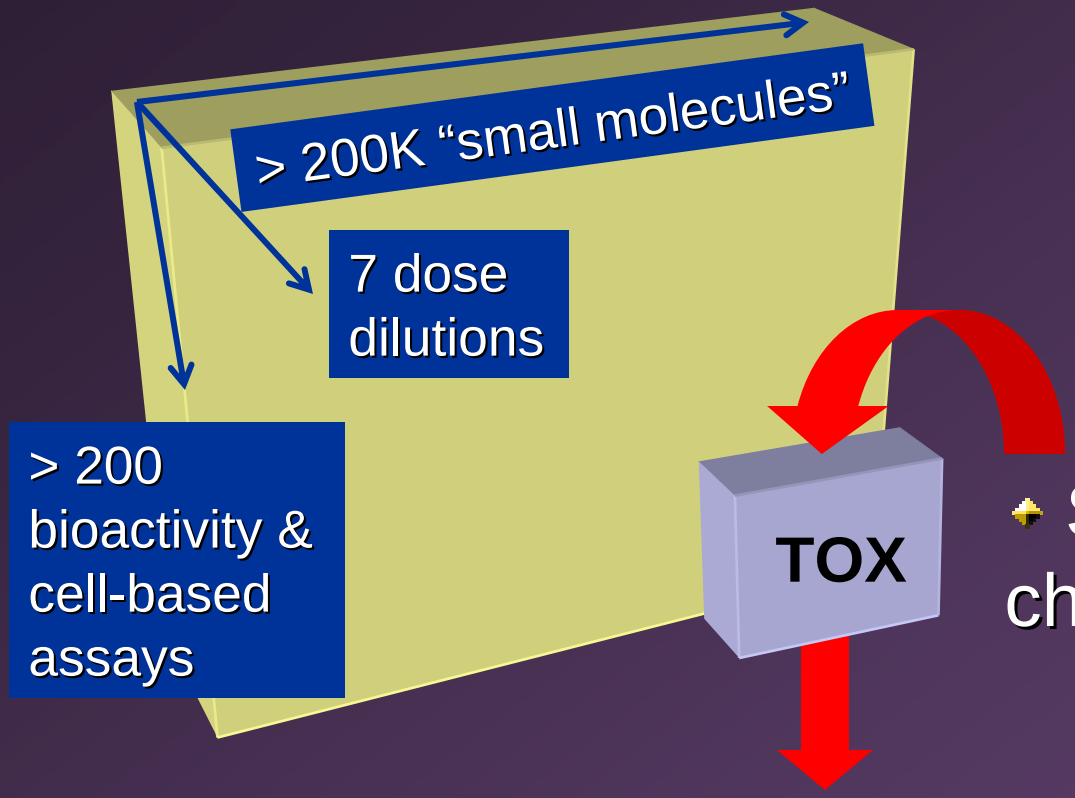


Approximate "proteomic" signature

NIH/NCGC Roadmap: nihroadmap.nih.gov

Small Molecules High-Throughput Screening Initiative

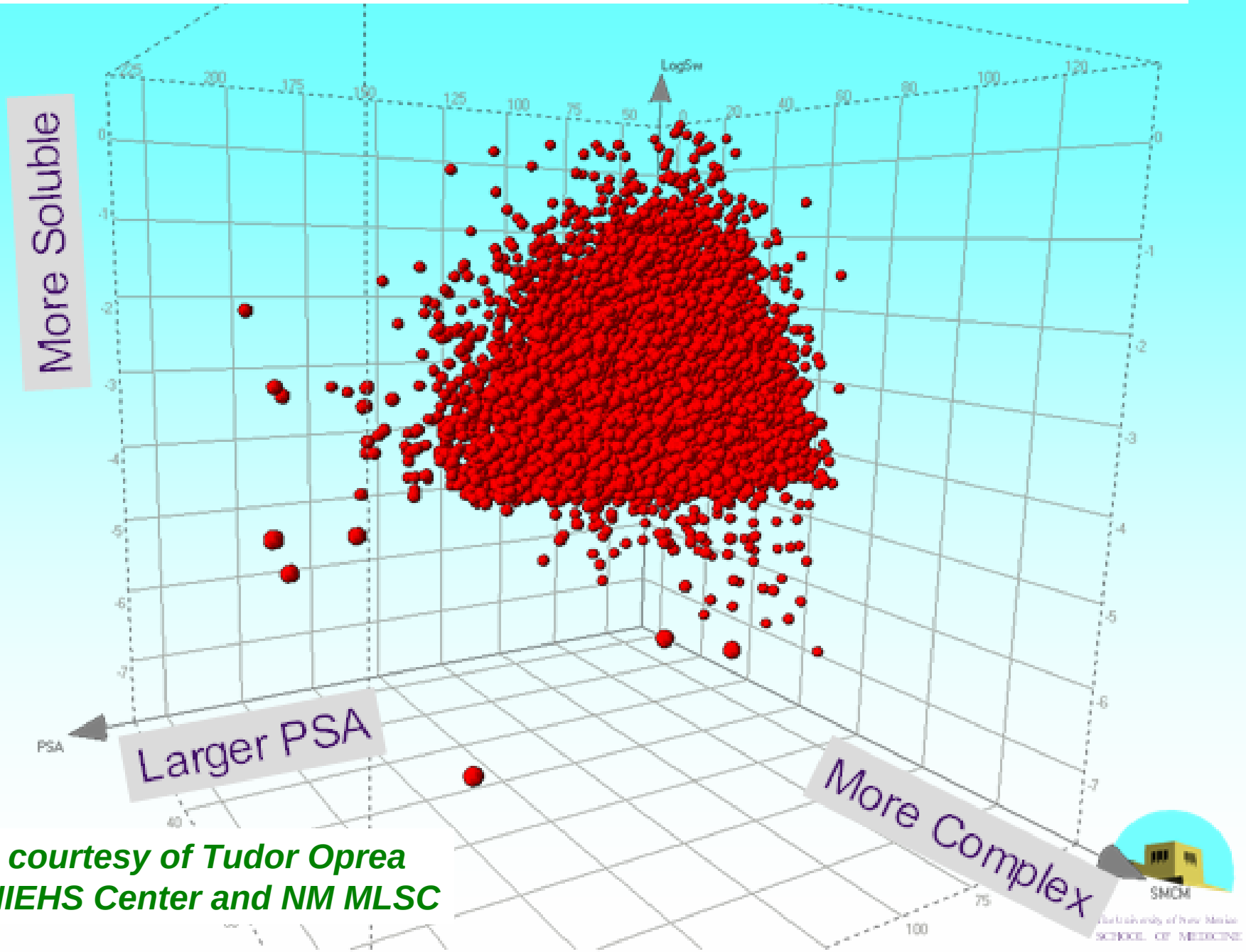
NIH Chemical Genomics Center



- Sample "toxicity" chemical space:
  - NTP chemicals
  - EPA pesticides
  - DSSTox
  - NCI/ChemNavigator

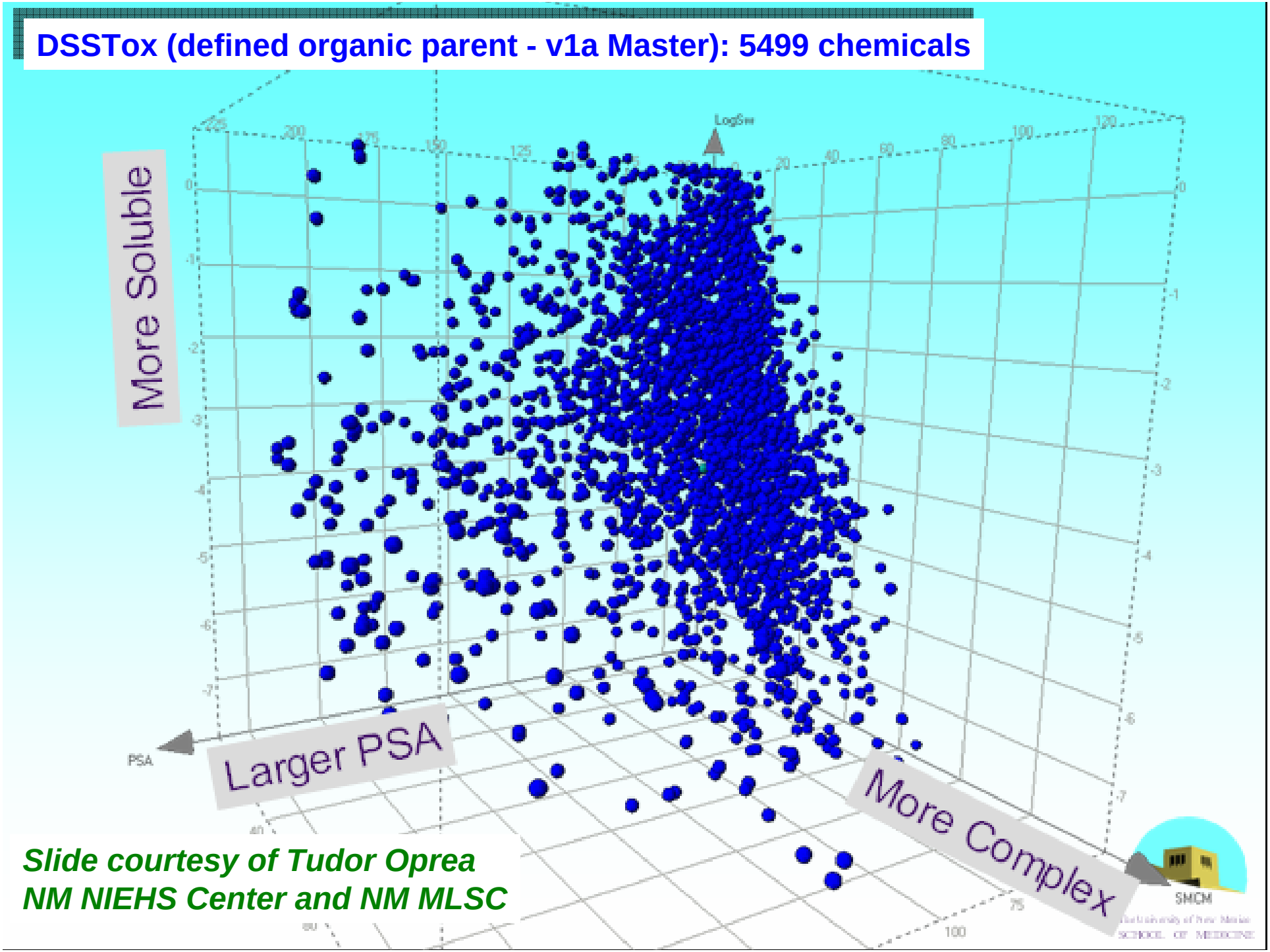
- Reference dataset of toxicity-related chemicals with structures & bioactivity profiles

NIH Molecular Roadmap (DPI) Small Molecules Repository: 65225 chemicals



Slide courtesy of Tudor Oprea
NM NIEHS Center and NM MLSC

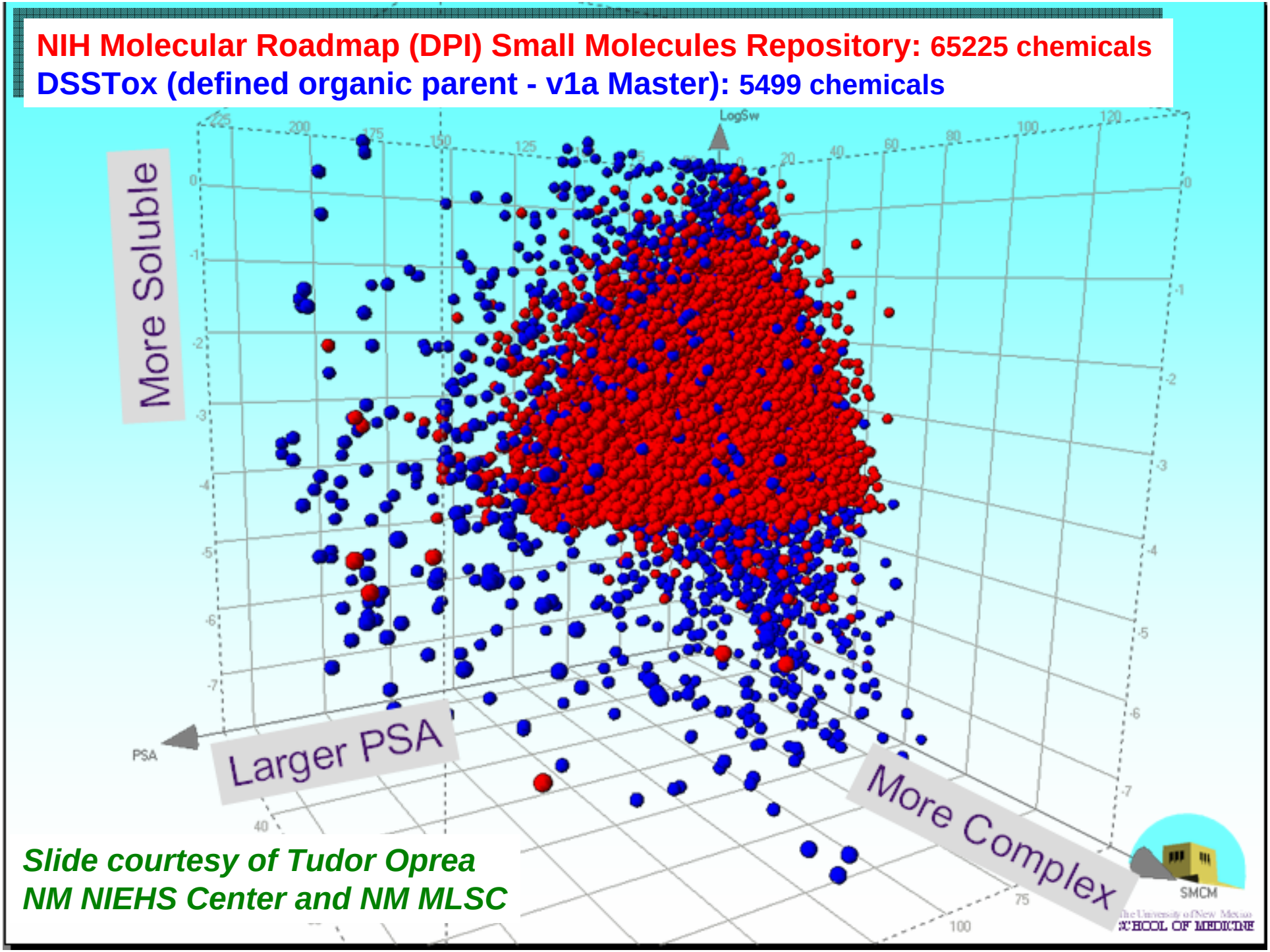
DSSTox (defined organic parent - v1a Master): 5499 chemicals



Slide courtesy of Tudor Oprea
NM NIEHS Center and NM MLSC

NIH Molecular Roadmap (DPI) Small Molecules Repository: 65225 chemicals

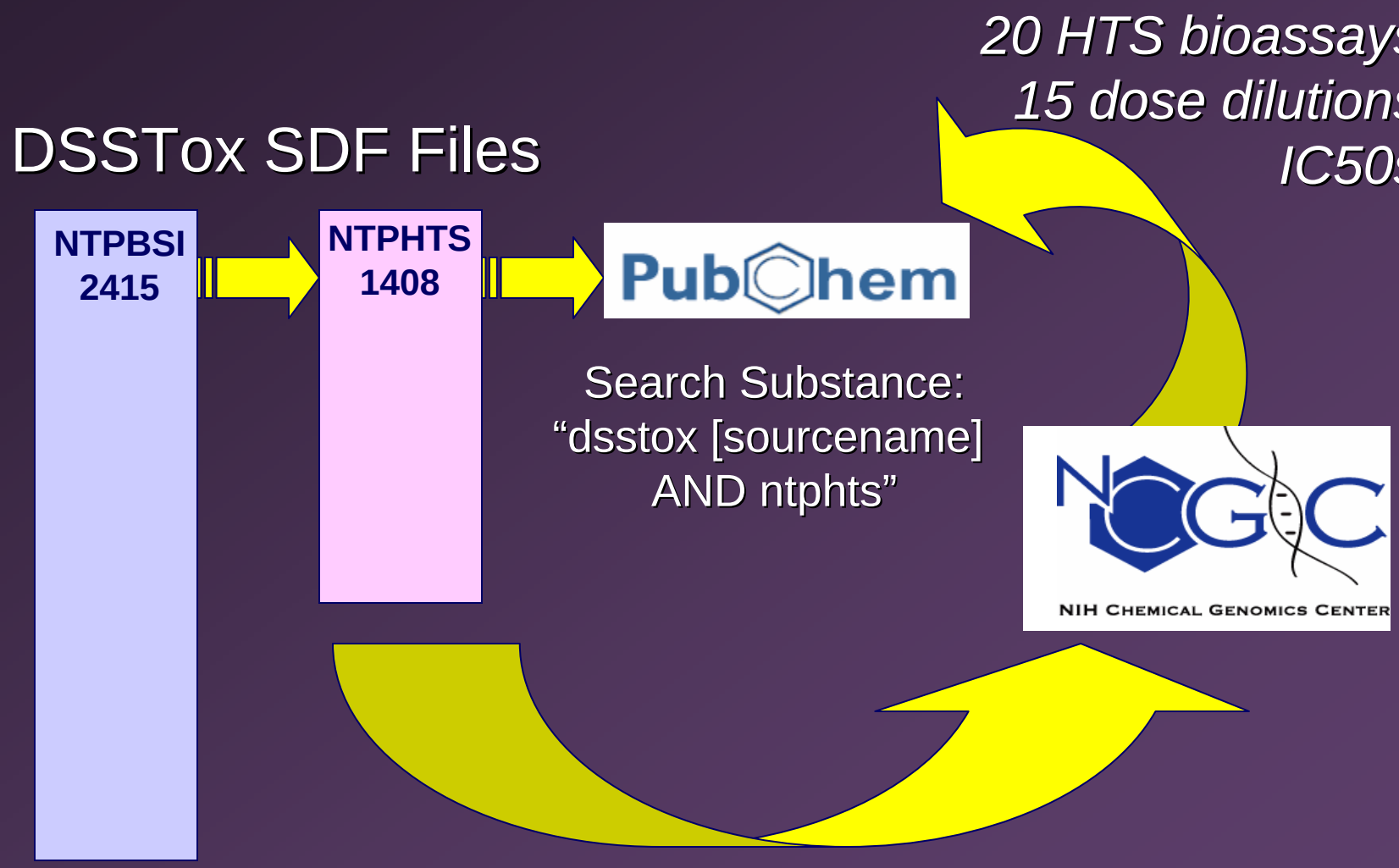
DSSTox (defined organic parent - v1a Master): 5499 chemicals



Slide courtesy of Tudor Oprea

NM NIEHS Center and NM MLSC

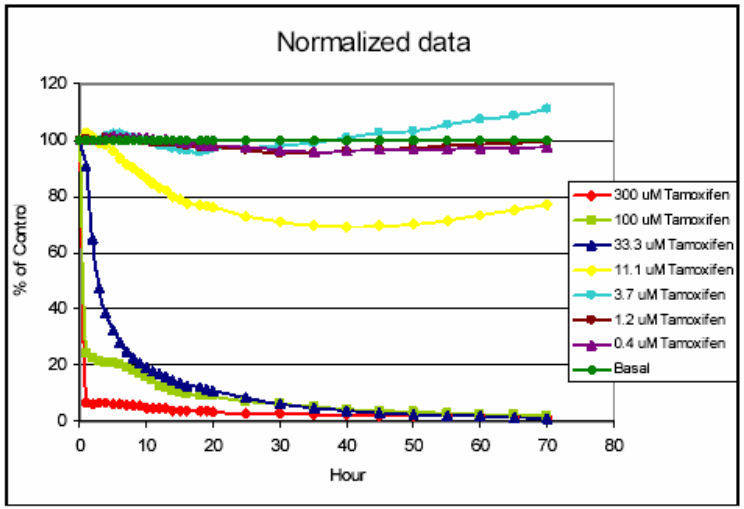
NTP High-Throughput Testing Program in Collaboration with NCGC

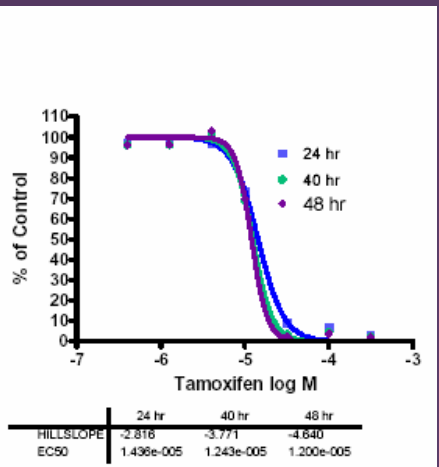


NTPHTS_NCGC Assay Results (1408)

- CellTiter-Glo luminescent assay for cell viability
- 4 human cell lines representing common tissue toxicities:
  - *Liver: HepG2 (human hepatocellular carcinoma)*
  - *Blood: Jurkat (Clone E6-1, human T cell leukemia)*
  - *Kidney: HEK293 (human embryonic kidney cells)*
  - *Nerve: SK-N-SH (human neuroblastoma)*
- 2 human primary cell lines:
  - *Human lung fibroblasts: MRC-5*
  - *Human foreskin fibroblasts: BJ*
- Compound profile in HepG2 cells

Time Course for Tamoxifen in HepG2 Cells





Courtesy of Chris Austin, NCGC

NCBI | PubChem | National Library of Medicine

HOME | SEARCH | SITE MAP | PubMed | Entrez | Structure | GenBank | PubChem | Help

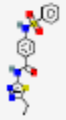
BioAssay Results

BioAssay ID (AID): 360
Source: NCGC
Name: Glucocerebrosidase

Source: “dsstox” AND “ntphts”

Total 48125 compounds found (48125 unique), 20 displayed: Next page

AC50, Hill Coefficient

Structure	PubChem SID	PubChem CID	Outcome	Activity Score	Submitter	Submission Date	Activity Direction	Activity Qualifier	Qualified AC50	Log of AC50	Hill Coefficient	Curve R2	Data Type	Compound Type
	4243169	3237927	Active	72	ncgc	19 Jan 2006	decreasing	=	6.06e-008	-7.22	0.87	1	qHTS	NIHSMR
	4264637	2210290	Active	71	ncgc	19 Jan 2006	decreasing	=	7e-008	-7.16	0.66	1	qHTS	NIHSMR

Full Concentration Response Curve

Compound QC	Curve Fit Model	Hill S0	Hill Sinf	Hill dS	Log AC50 Std Error	Curve Chi2f	Excluded Points	Number of Points	Activity at 4.925nM (%)	Activity at 24.623nM (%)	Activity at 0.123uM (%)	Activity at 0.615uM (%)	Activity at 3.077uM (%)	Activity at 15.386uM (%)	Activity at 0.077mM (%)
QC'd by DPI	4pHill (AC50,n,S0,Sinf)	-1.23	-100.1	98.92	0.02	0.5	{}	7	-11.3	-31.9	-65.8	-88.4	-96.4	-99.6	-100.2
QC'd by DPI	4pHill (AC50,n,S0,Sinf)	11.41	-107.2	118.6	0.1	2.04	{}	7	-5.5	-30.2	-56.9	-84.4	-100	-103.7	-105

EPA ToxCast Program

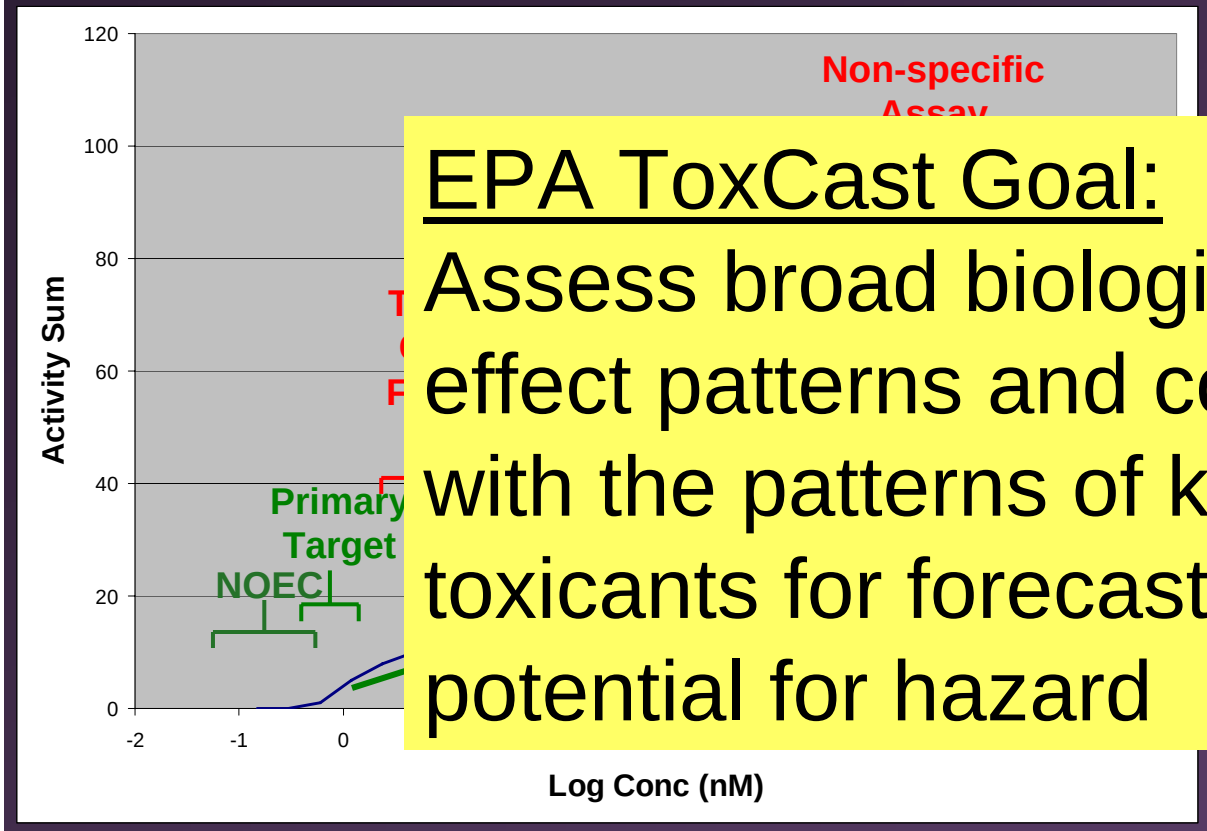
- HTS screening of 300-400 pesticide actives in > 100 selected bioassays; 1408 total compounds for NCGC screening
  - *bioactivity*
  - *cell/tissue-based*
  - *high content assays (e.g. c elegans)*
- ToxRefDB - High quality reference data for registered pesticides
  - *create relational database (tailor ToxML schema)*
  - *extract data from EPA Registered Pesticide DERs*
- DSSTox structure annotation, overlaps with toxicity databases, chemical selection criteria
- ACToR integrated data warehouse & analysis system to support prediction modeling efforts

Assay Coverage –
Key Mechanisms / Toxicities

- Cell cycle, apoptosis, DNA recombination and repair
- Transporters, channels, membrane receptors
- Signal transduction pathways
- Nuclear receptor mediated pathways
- Oxidative Stress

- Genotoxic and non-genotoxic carcinogenicity
- Developmental and reproductive toxicity
- Developmental neurotoxicity and immunotoxicity

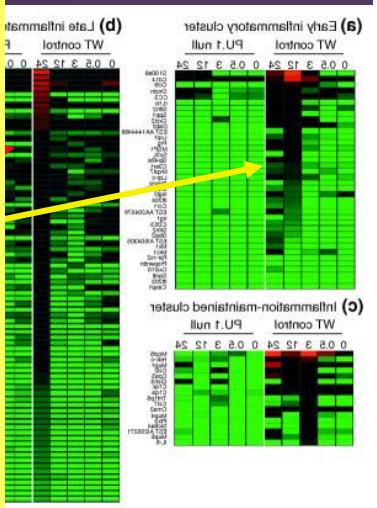
Bioactivity Profiling of Pharmaceuticals vs Environmental Chemicals



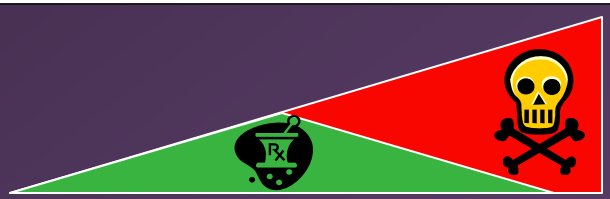
EPA ToxCast Goal:

Assess broad biological effect patterns and correlate with the patterns of known toxicants for forecasting potential for hazard

HTS Target Panels

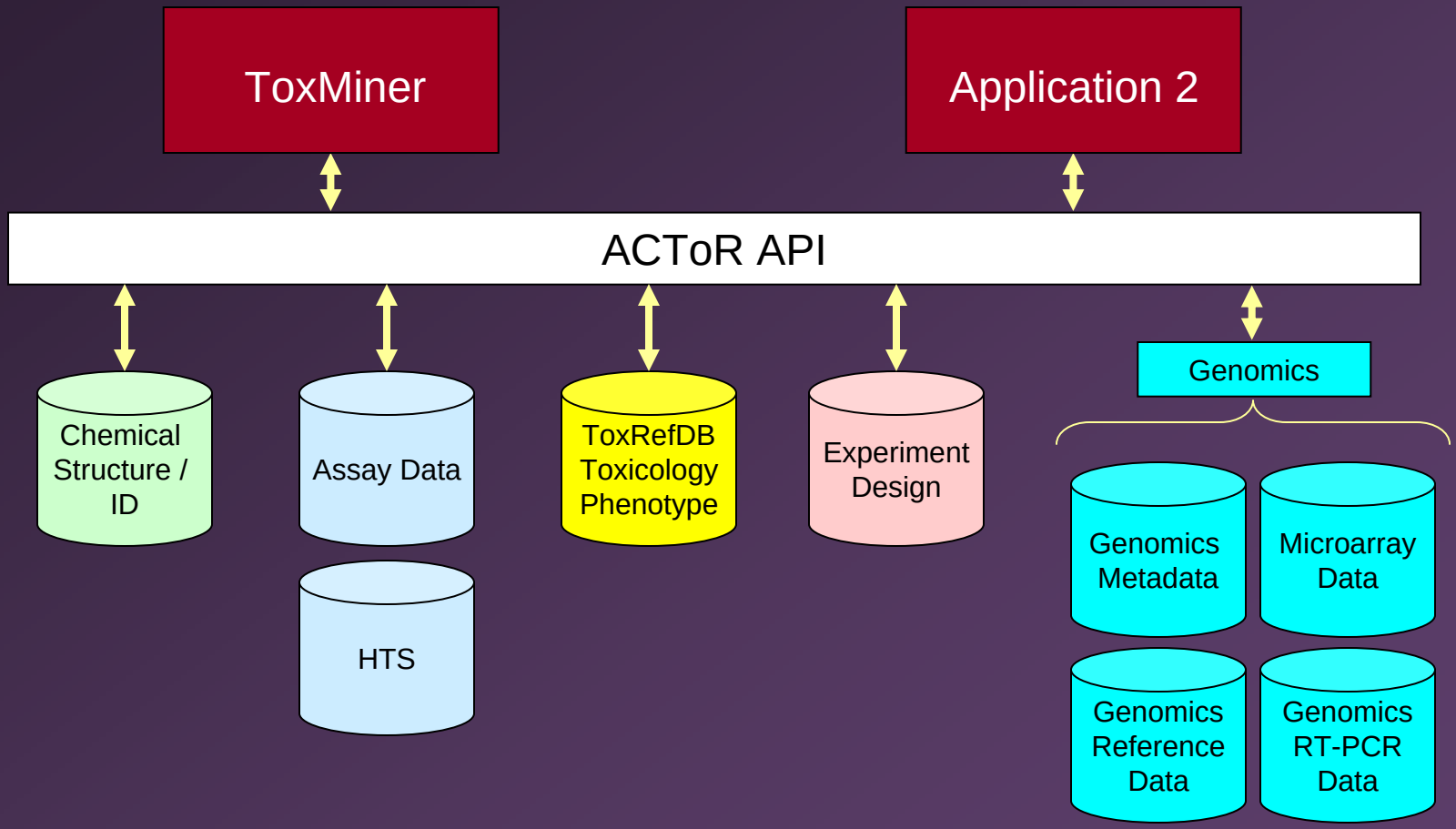


Heat map profiles of activity in different assays (columns) arranged by compound (rows). Inactive (green) to highly active (red).



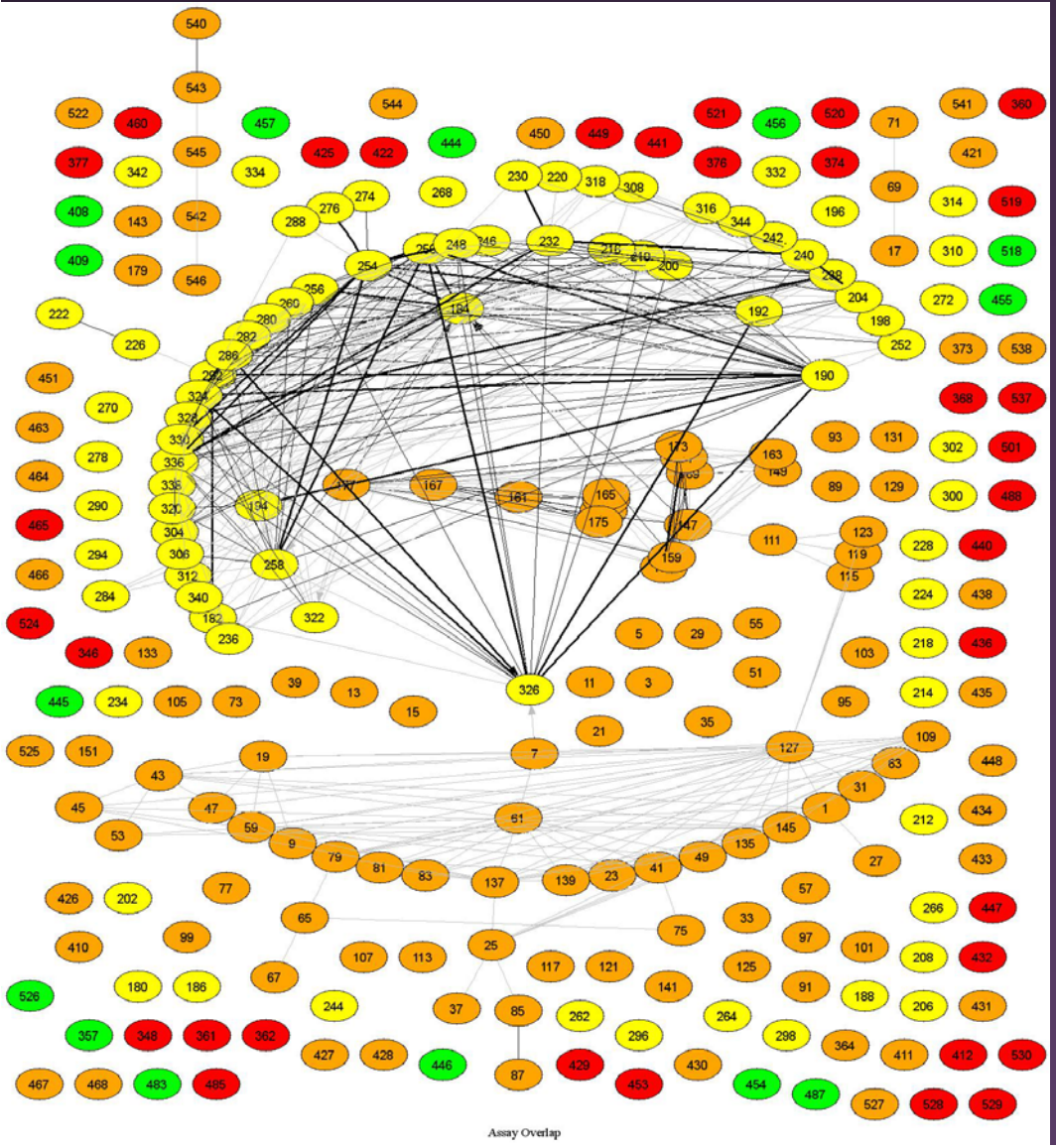
Slide courtesy of Dr. Keith Houck, EPA National Center for Computational Toxicology

ACToR
Aggregated Computational Toxicology Resource



Courtesy of Dr. Richard Judson, EPA National Center for Computational Toxicology

Example Correlation Map From PubChem (NIH)



252 assays from diverse sources

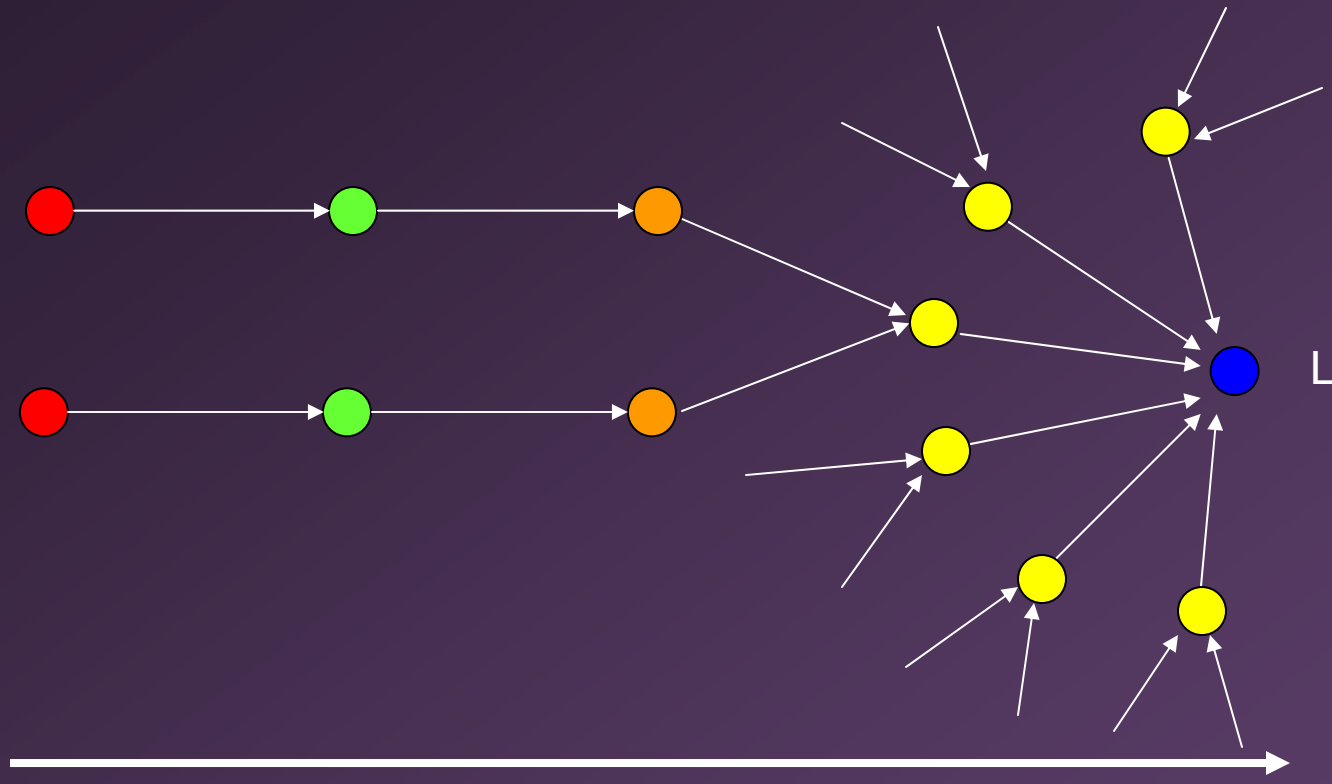
Line represent strong correlation from results of common chemicals

ToxCast will far exceed PubChem in number of assays spanning a common set of chemicals

- Direct Molecular Interaction
- Molecular Pathway
- Cellular Process
- Tissue/Organ Effect
- Organism Endpoint

Courtesy of Dr. Richard Judson, EPA National Center for Computational Toxicology

Predictive Power Increases as We Move Upstream



- Direct Molecular Interaction
- Molecular Pathway
- Cellular Process
- Tissue/Organ Effect
- Organism Endpoint

Courtesy of Dr. Richard Judson, EPA
National Center for Computational Toxicology

EPA ToxCast
ToxRefDB - Toxicity Reference Database:

Source:	EPA's Office of Pesticide Programs (OPP)
Format:	Data Evaluation Record (DER)
Chemical:	Conventional Pesticide Active Ingredients (~800)
Purity:	Technical Grade (>90%)
Dosing:	Primarily Orally Administered (based on availability and use pattern of pesticide)
Study Type:	Subchronic Toxicity (Rodents and Non-Rodents)
	Prenatal Developmental Toxicity
	Reproduction and Fertility Effects (2-generation)
	Chronic Toxicity (rat, mouse, and dog)
	Carcinogenicity (rat and mouse)
	Developmental Neurotoxicity
	Immunotoxicity

Targeted Toxicological Data Collection

Data Collection Results

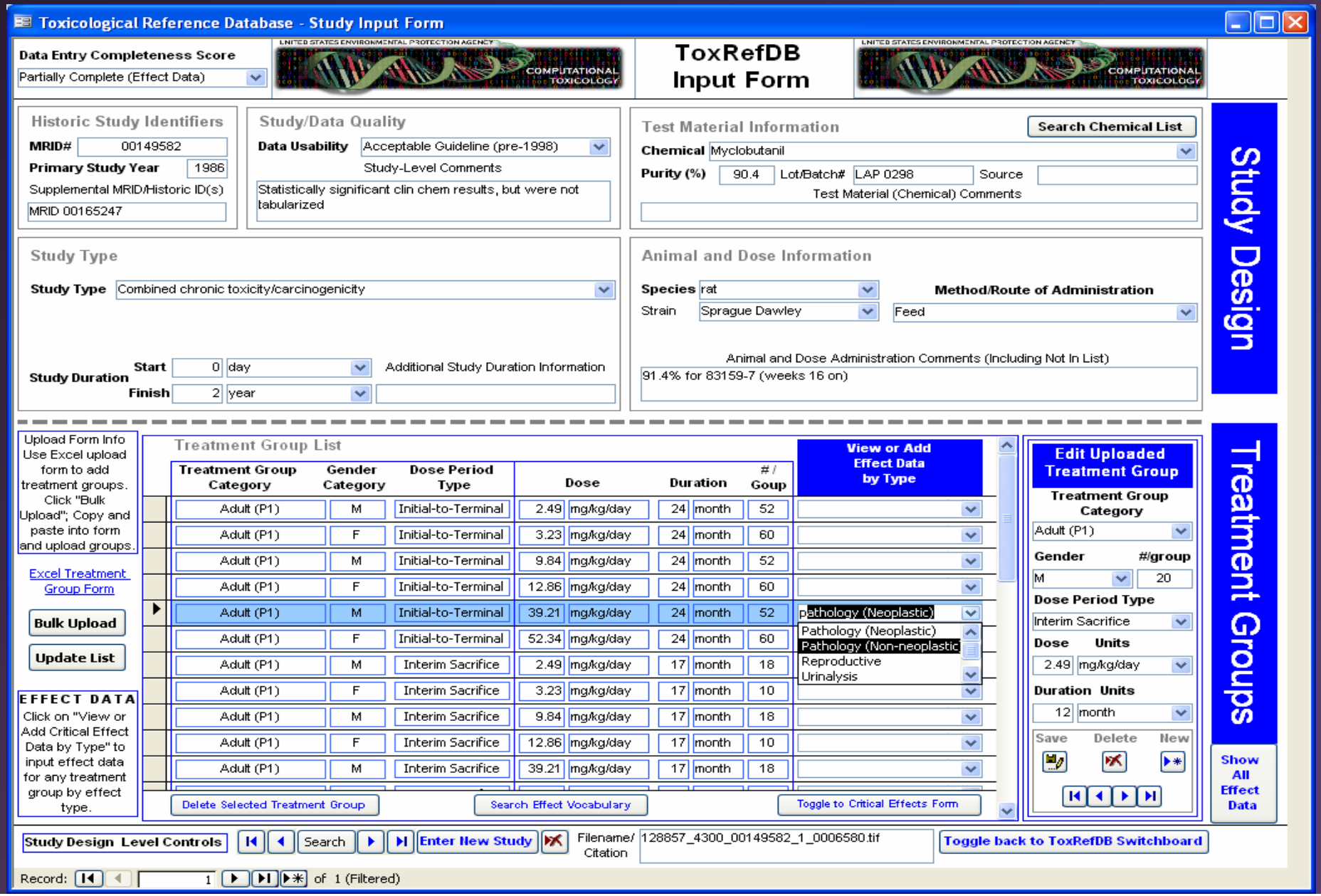
> 4000 DER (2500 studies)
for over 400 pesticides

Toxicological Schema and Lexicon Development

- ToxML compatibility and interoperability
- Standardized fields and vocabulary
  - Study Type (OPPTS/OECD Test Guidelines)
  - Data Usability (Data Quality) Code
  - Animal Info (Species, Strain, Sex Category)
  - Treatment Group Category (Adult, Offspring, etc.)
  - Endpoint Category (Systemic, Maternal, etc.)
  - Effect Descriptors (Type, Target, and Description Vocabulary)

Courtesy of Matt Martin, EPA National Center for Computational Toxicology

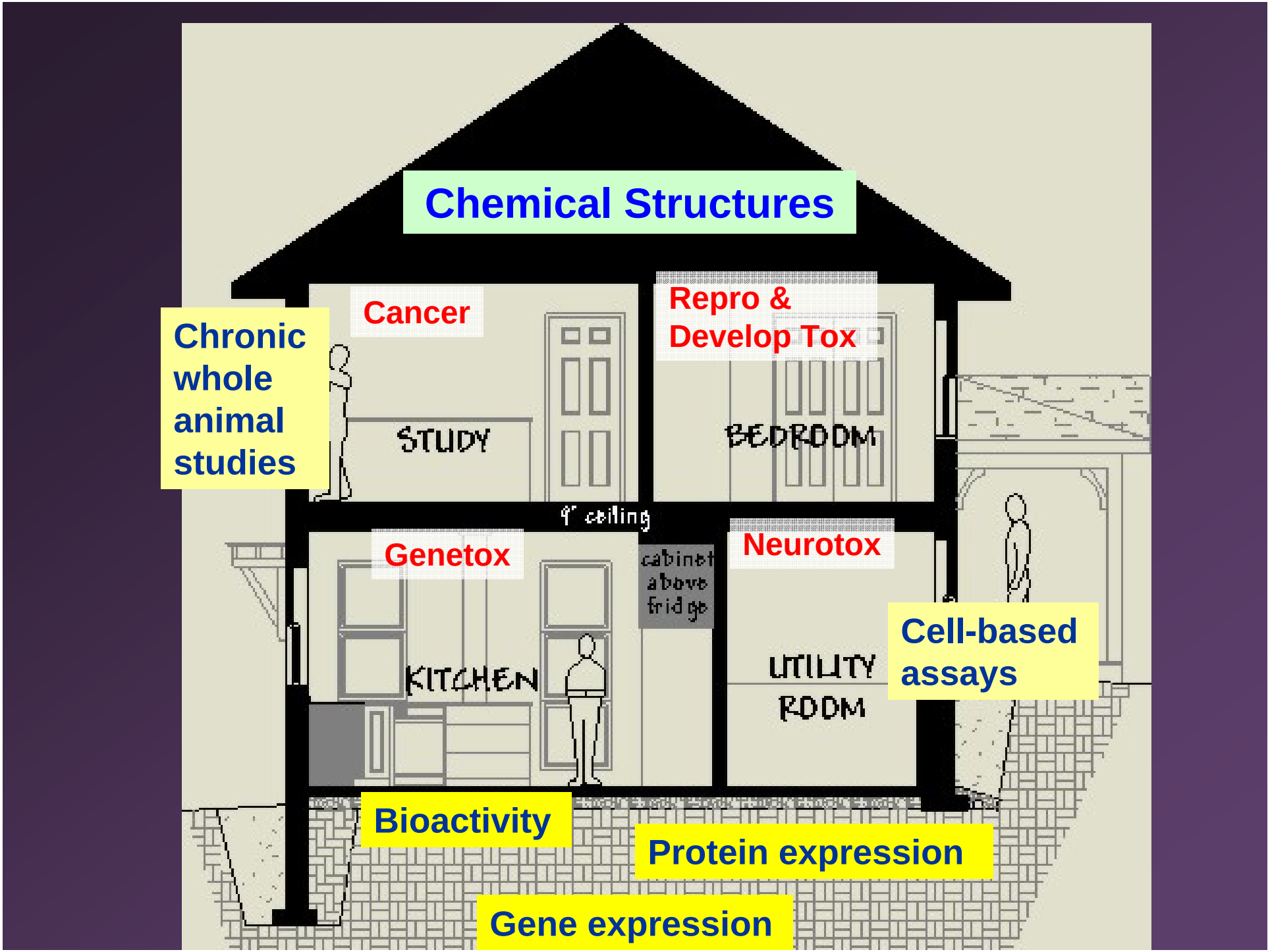
EPA ToxCast Toxicity Reference Database

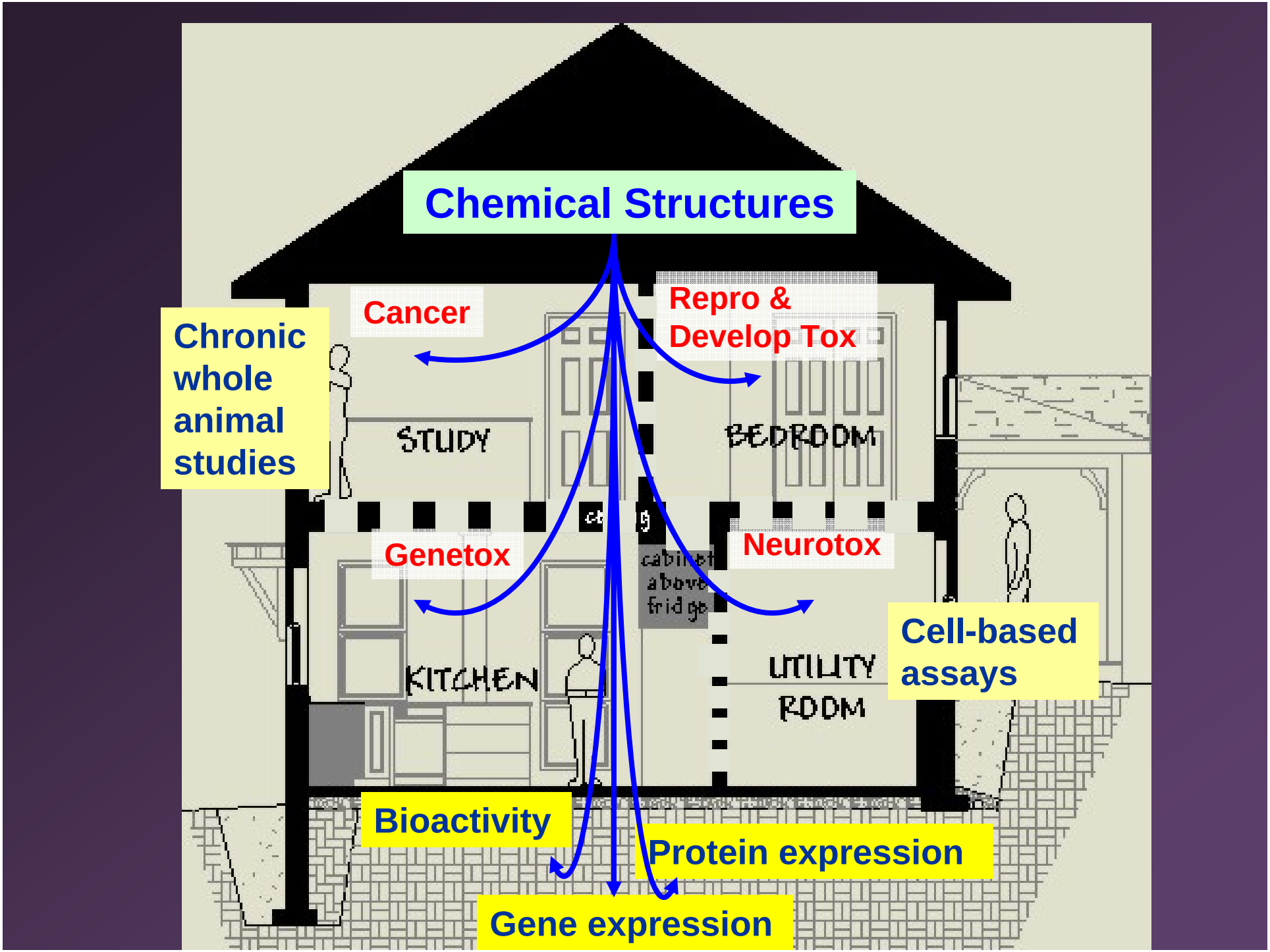


ToxCast Toxicity Reference Database

Summary of 53 Chronic/Cancer Rat/Mouse Studies Entered into ToxRefDB			
Total Effects Entered		**1452**	
Unique Effects		**317**	
Effect Type		**Endpoint Summary**	
Body Weight	166	Studies achieve LOAEL	40
Organ Weight	205	**Endpoint Dose Range (mg/kg/day)**	
Clinical Chemistry	85	<=10	9
Hematology	116	>10 and <=100	20
Non-neoplastic Pathology	333	>100 and <=1000	10
Clinical Signs	66	>1000	1
Effect Target		**Effects at Dose Range (mg/kg/day)**	
Liver	298	<=10	274
Kidney	70	>10 and <=100	536
Testes	49	>100 and <=1000	561
Thryoid Gland	14	>1000	49

Courtesy of Matt Martin, EPA National Center for Computational Toxicology


Chemical Structures
Chronic whole animal studies
Cancer
Repro & Develop Tox
STUDY
BEDROOM
9' ceiling
Genetox
Neurotox
cabinet above fridge
KITCHEN
UTILITY ROOM
Cell-based assays
Bioactivity
Protein expression
Gene expression


Chemical Structures
Chronic whole animal studies
Cancer
Repro & Develop Tox
Genetox
Neurotox
Cell-based assays
Bioactivity
Protein expression
Gene expression
STUDY
BEDROOM
KITCHEN
UTILITY ROOM
cabinet above fridge

Acknowledgments

- Maritja Wolf, Jamie Burch, ClarLynda Williams – EPA/NCCT/DSSTox
- Ray Tice, Cynthia Smith, Beth Bowden, Brad Collins – NIEHS/NTP
- Robert Kavlock (Director, NCCT), David Dix (ToxRefDB, HTS, Genomics), Keith Houck (HTS), Matt Martin (ToxRefDB), Richard Judson (ACToR) – EPA/NCCT/ToxCast
- Chihae Yang – Leadscope
- Steve Bryant, Jane Tseng, Yanli Wang – PubChem Project
- Lois Gold – Carcinogenic Potency Project
- Tudor Oprea – Univ. of New Mexico, School of Medicine
- Chris Austin, Ajit Jadhav (NIH/NCGC); Doug Livingston (DPI-SMR)