United States Environmental Protection Agency

Dalton State College, GA, April 20, 2017

The art and practice of structure-activity modeling at the chemistry-toxicology interface

Ann Richard

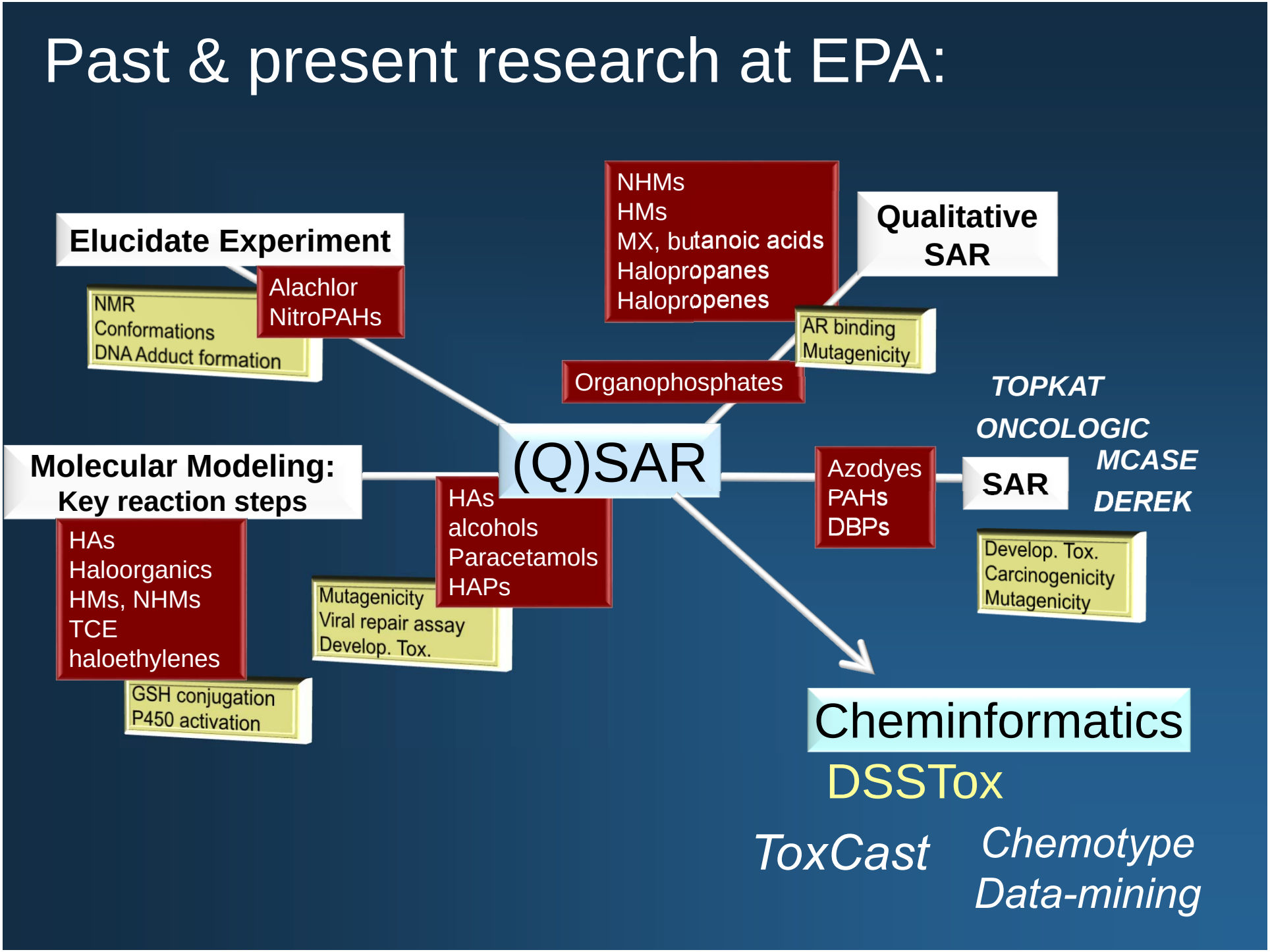
National Center for Computational Toxicology

Office of Research & Development

richard.ann@epa.gov

Office of Research and Development
National Center for Computational Toxicology

Past & present research at EPA:



Part I:
Computational chemistry & mechanism-based structure modeling

Focused application
Narrowly defined questions

United States Environmental Protection Agency

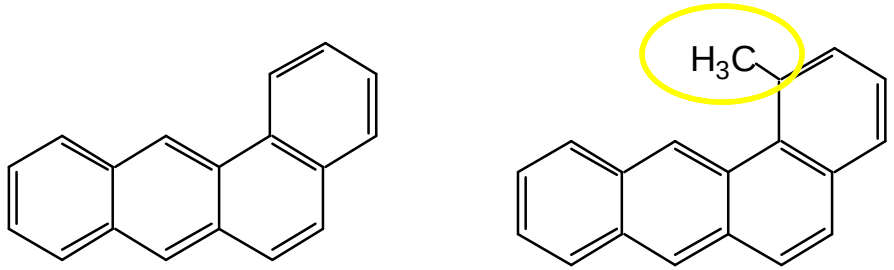
Structure-**A**ctivity **R**elationship (**SAR**) postulate:

similar molecules have similar activities

(except when they don't!)

Structure-Activity Relationships (SAR)

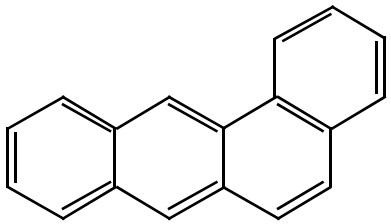
Activity = f (Structure)



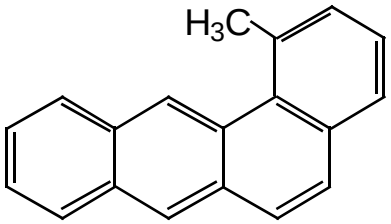
active inactive

Methyl group leads to loss of carcinogenic activity

SAR Generalization



active



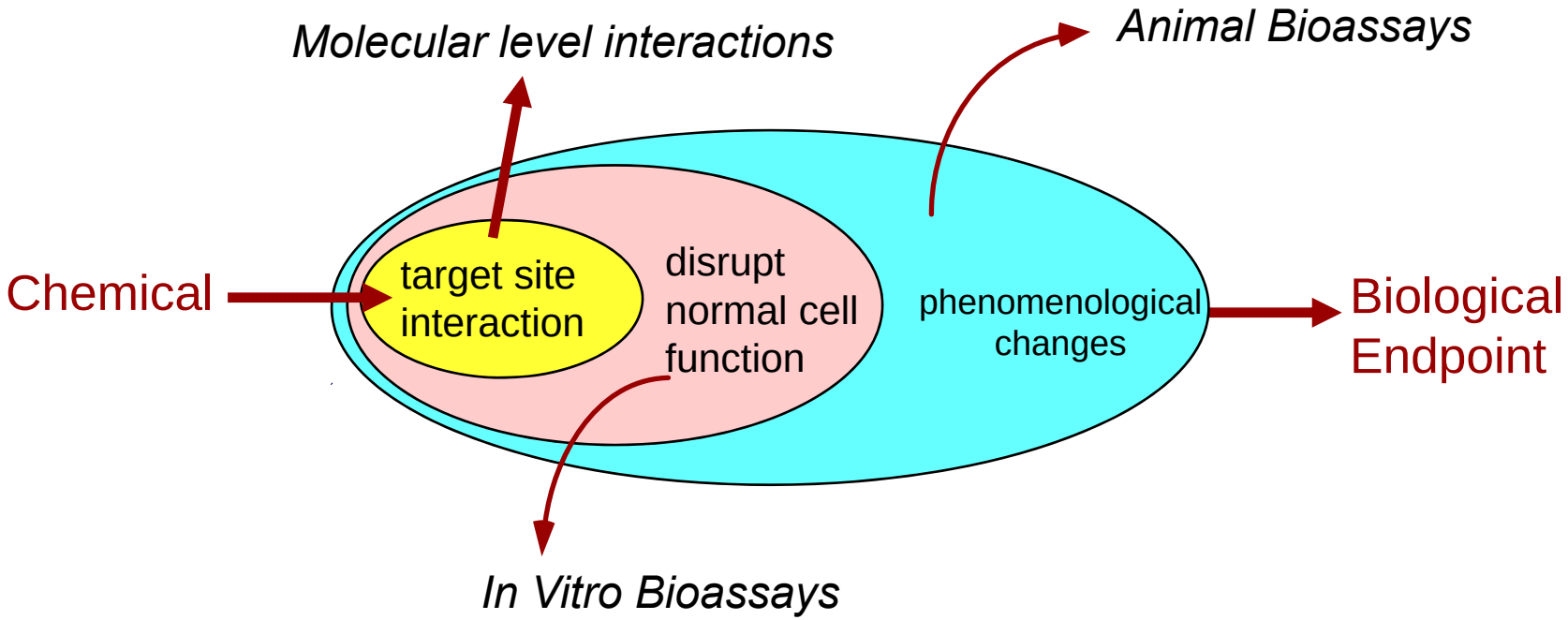
inactive

Statistical association

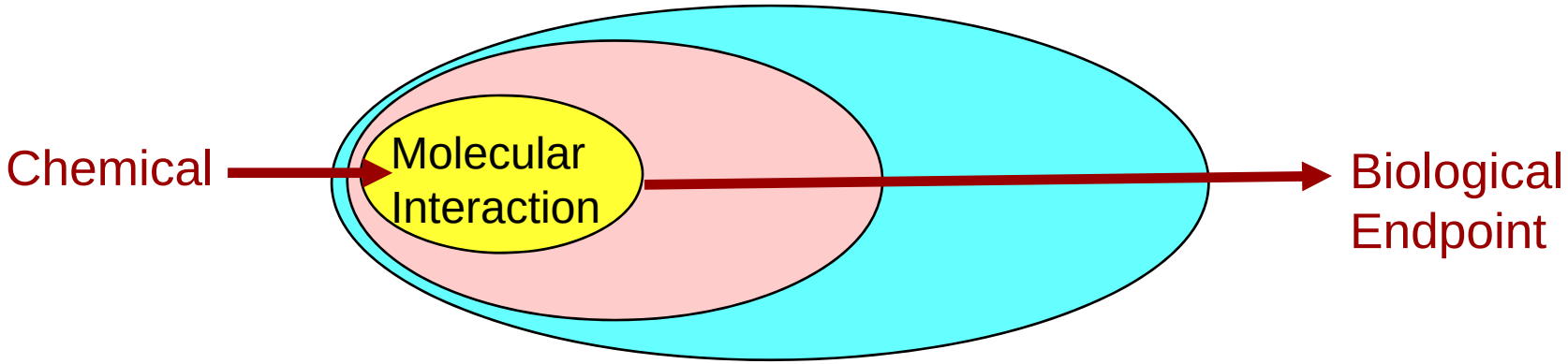
Mechanistic hypothesis

Class	→	PAHs
activating feature	→	bay region
modulating feature	→	steric hindrance

Mechanisms of Toxicity



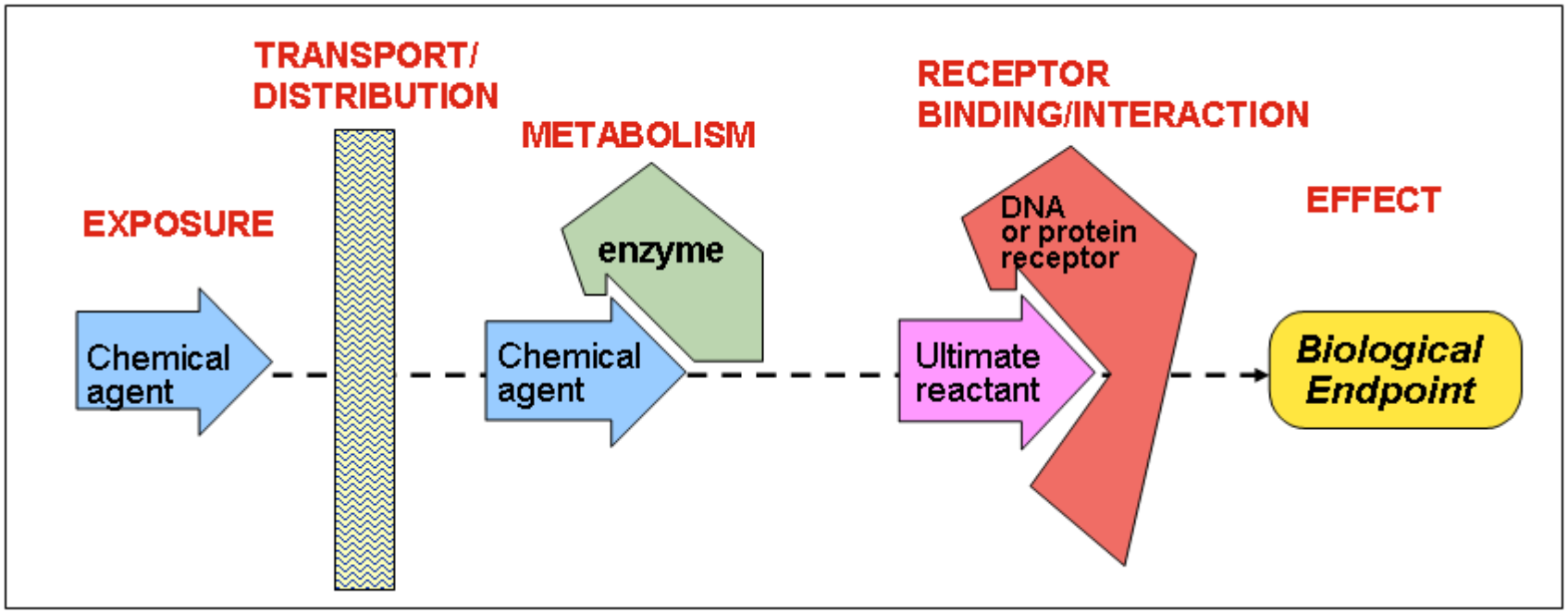
Structure-Activity Relationships



Similar chemicals
Relative properties
Common mechanism of action

Typical (Q)SAR Paradigm

Activity = Constant * $Prob_{site}$ * $Prob_{rxn}$

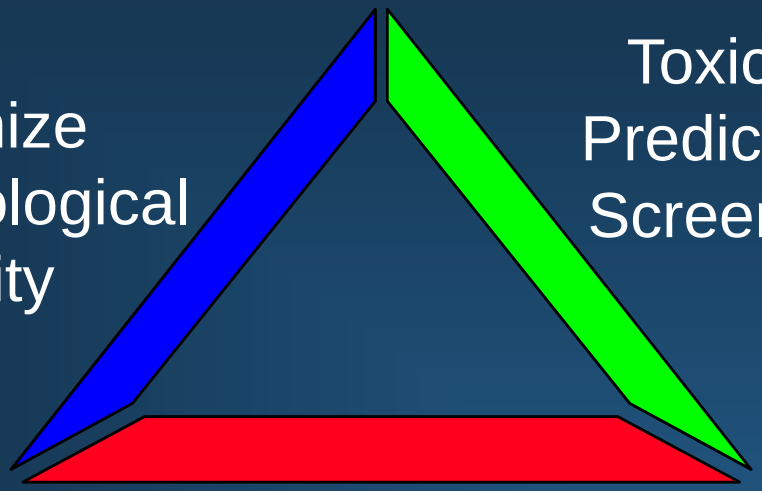


vapor pressure
Solubility
Henry's const

log P(octanol/water)
acidity

electronic/ steric
3D properties
reactivity indicators
interaction energies

SAR Application



Maximize Pharmacological Activity

Toxicity Prediction/ Screening

Minimize Toxicity

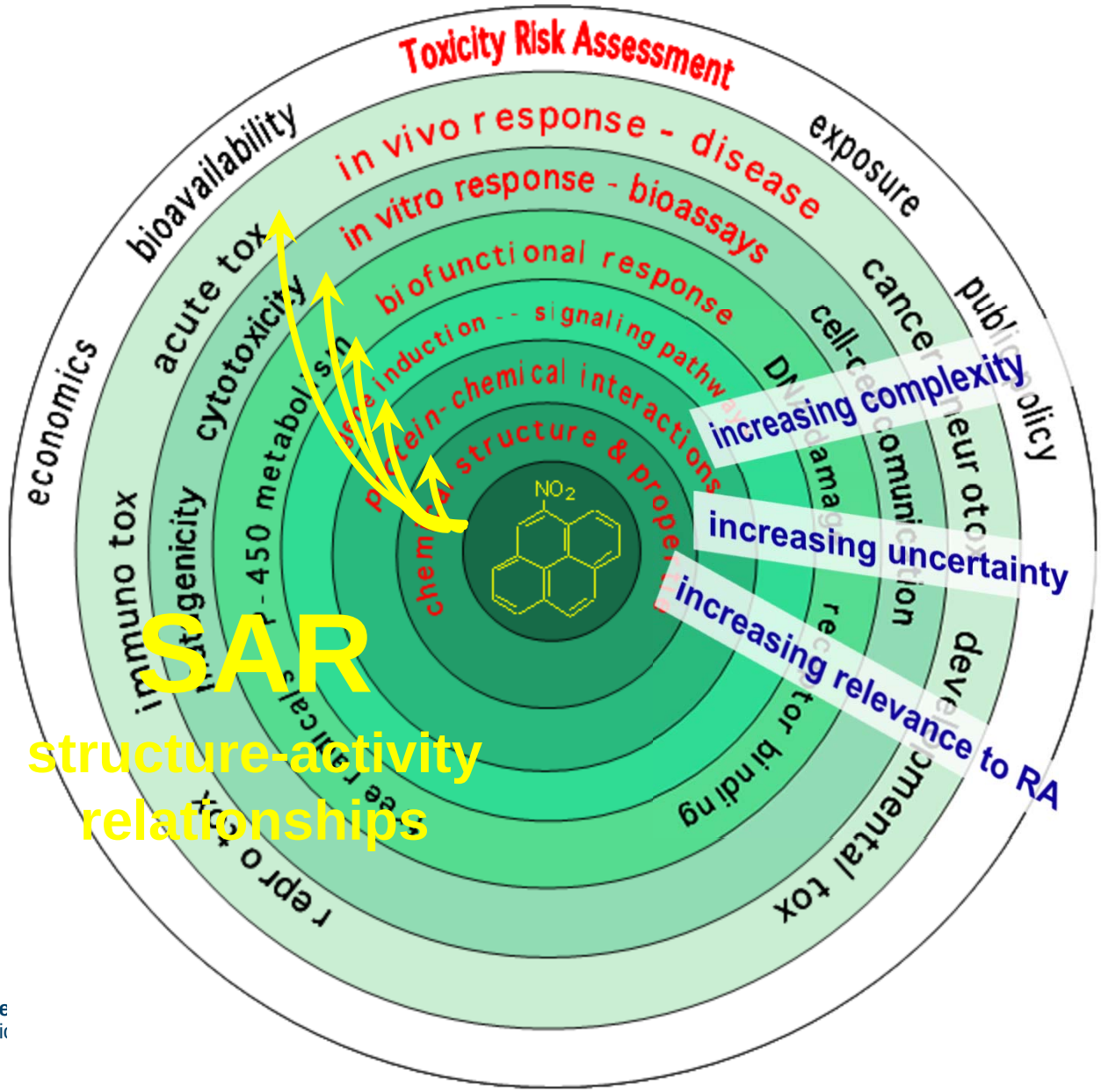
Drug Design

Environmental Protection

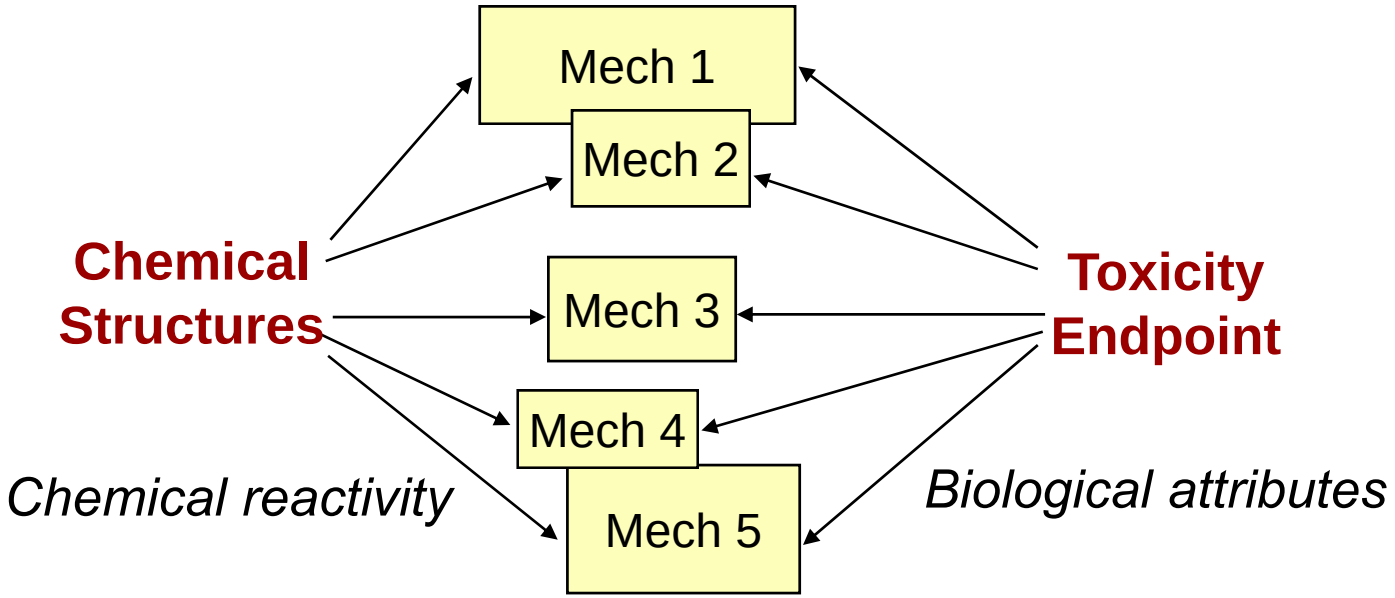
- Single therapeutic target
- Drug-like chemicals
- Some toxicity anticipated

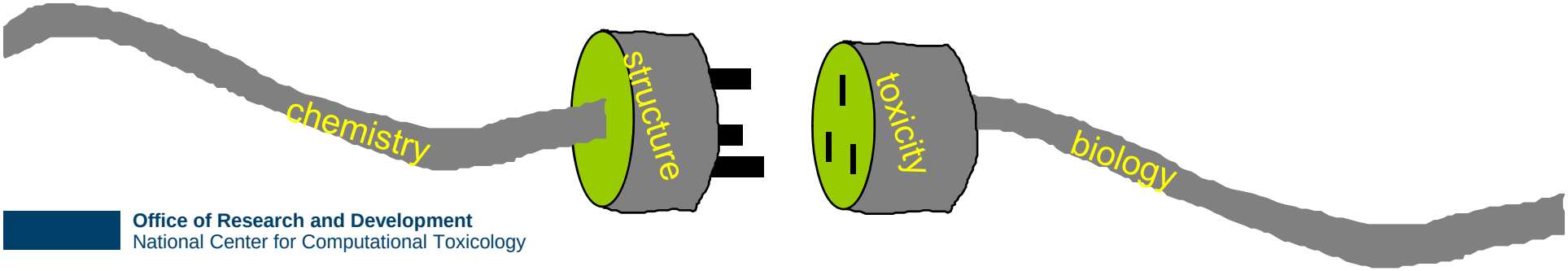
- Multiple unknown targets
- Diverse structures
- Human and eco endpoints

Toxicity Prediction Problem



Global vs. Local SAR models





Classical QSAR

“We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.”

- Isaac Newton

Quantitative Structure-Toxicity Relationships for a Series of Alcohols in a Mammalian Viral Host Cell Reactivation Assay

Benane SG, Richard AM, Blackman CF, Lytle CD (1993) InVitroToxicol 6(4):267-277.

Alcohol	Formula	HCR-inhibiting Concentration [mM]	LogP(o/w)
methanol	CH3OH	230.00	-0.77
ethanol	CH3CH2OH	100.00	-0.31
1-propanol	CH3CH2CH2OH	19.00	0.25
1-butanol	CH3(CH2)3OH	8.00	0.8
1-pentanol	CH3(CH2)4OH	2.20	1.56
1-hexanol	CH3(CH2)5OH	0.44	2.03
1-heptanol	CH3(CH2)6OH	0.19	2.41
1-octanol	CH3(CH2)7OH	0.08	2.97
2-propanol	CH3CH(OH)CH3	44.00	0.05
2-butanol	CH3CH2CH(OH)CH3	12.00	0.61
2-methyl-1-propanol	CH3CH(CH3)CH2OH	9.60	0.76

Congeneric series of chemicals

Reasonable to assume common mechanism of action

LogP(o/w) major determinant in many QSARs

Does inhibition of DNA repair for a series of alcohols exhibit typical "narcosis" activity, or a structure-activity cutoff (C7) as reported for permeability of bacterial membranes?

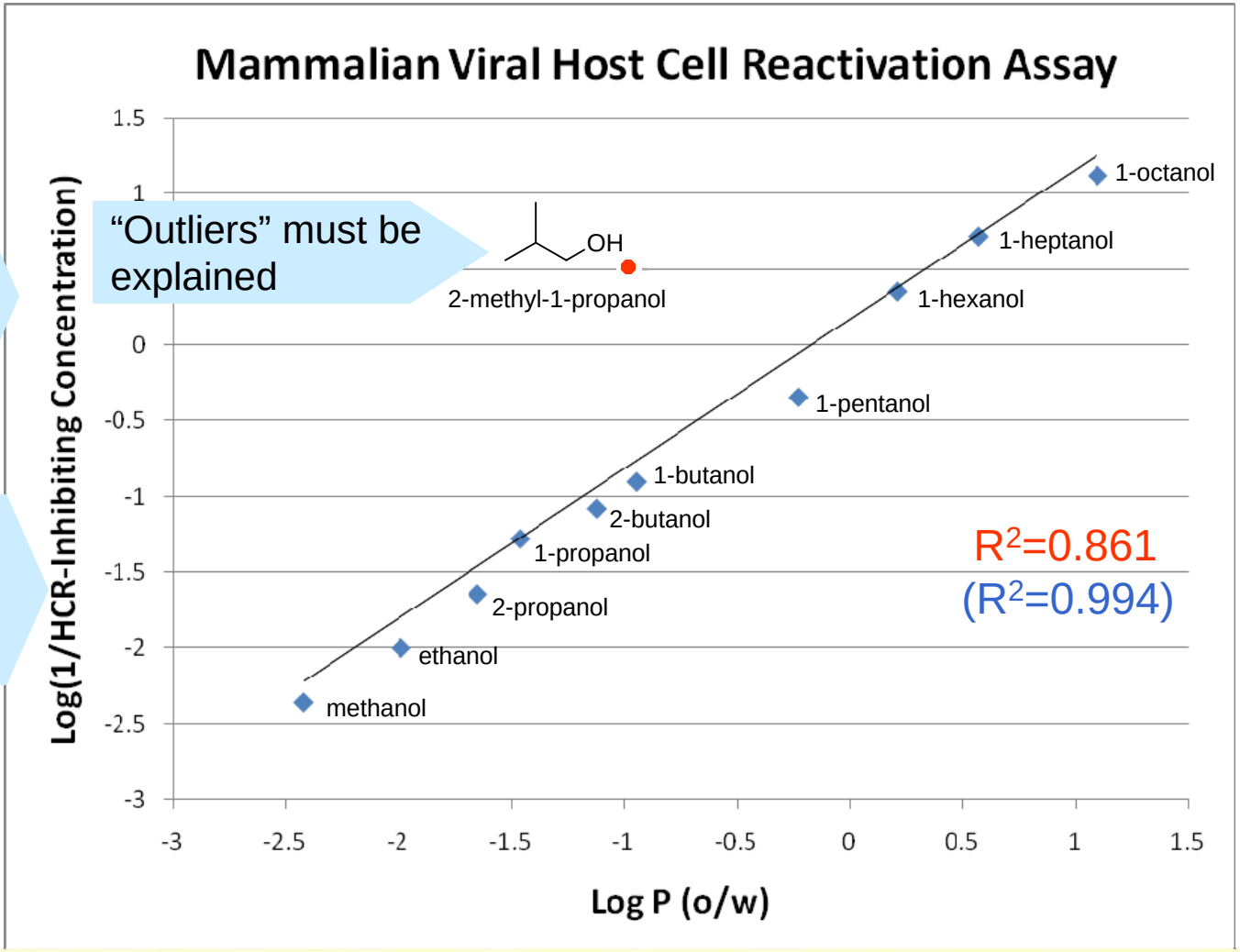
Quantitative Structure-Toxicity Relationships for a Series of Alcohols in a Mammalian Viral Host Cell Reactivation Assay

Benane SG, Richard AM, Blackman CF, Lytle CD (1993) InVitroToxicol 6(4):267-277.

Linear QSAR relates to log [1/Activity(molar concentration]

"Outliers" must be explained

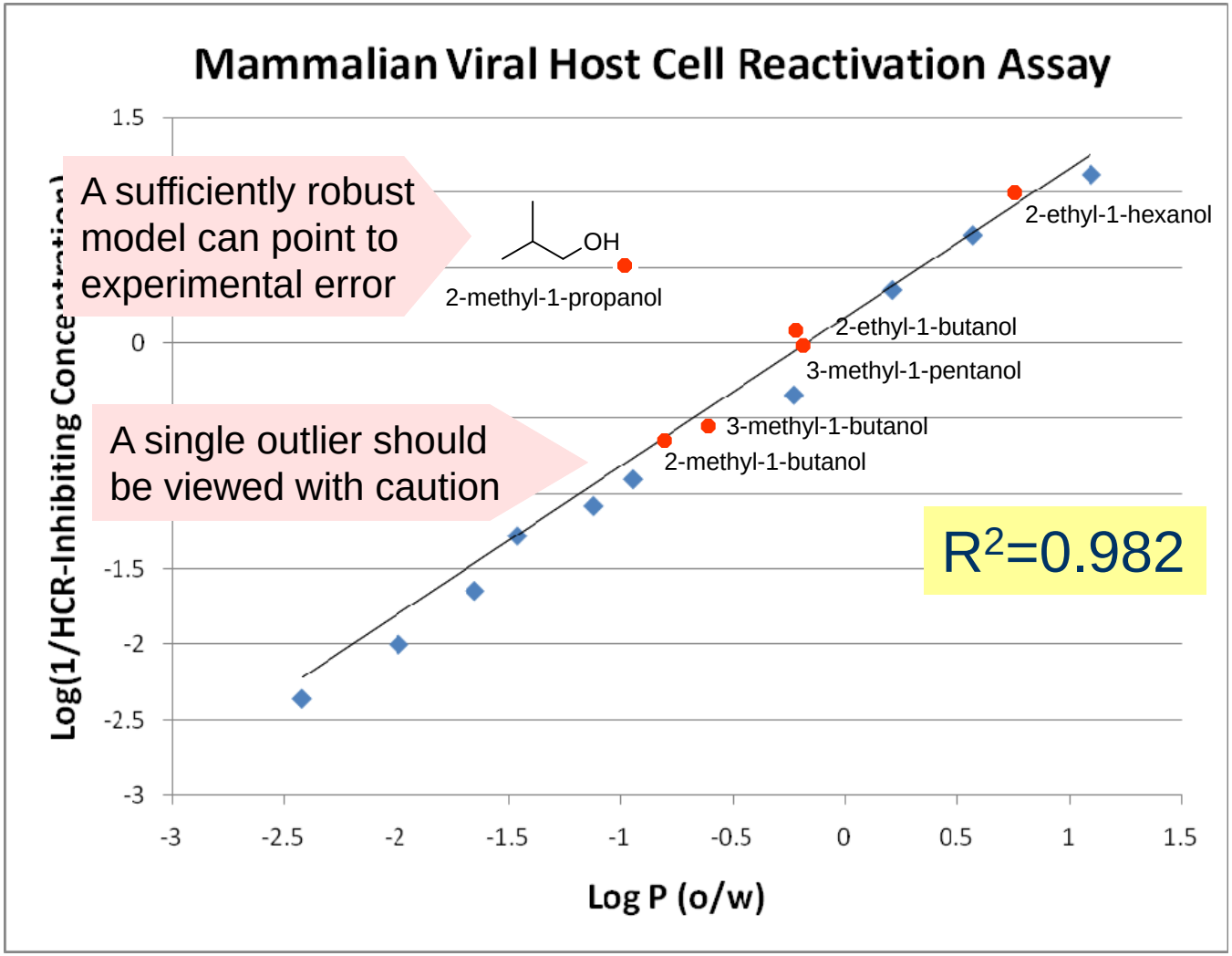
Activity spans several orders of magnitude



Does inhibition of DNA repair for a series of alcohols exhibit typical "narcosis" activity, or a structure-activity cutoff (C7) as reported for permeability of bacterial membranes?

Quantitative Structure-Toxicity Relationships for a Series of Alcohols in a Mammalian Viral Host Cell Reactivation Assay

Benane SG, Richard AM, Blackman CF, Lytle CD (1993) InVitroToxicol 6(4):267-277.



Do branched alcohols (similar to valproic acid) exceed "baseline" activity due to, e.g., active transport? → Test series of branched alcohols.

If only we could combine apples & oranges…

"Simply stated, it is sagacious to eschew obfuscation."

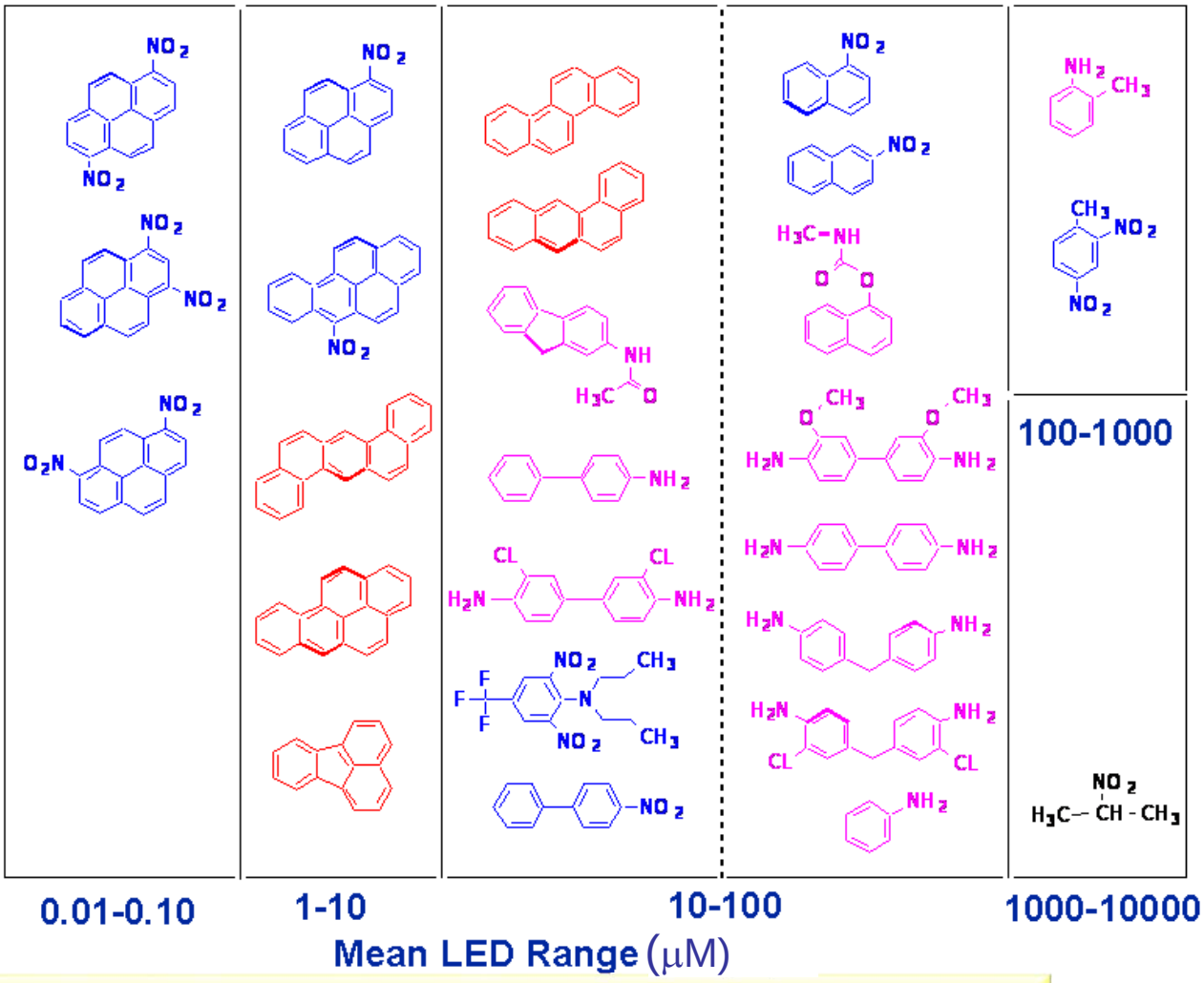
- Norman R. Augustine

QSARs of mutagens & carcinogens: Case study illustrating problems in the construction of models for noncongeneric chemicals

Benigni R, Richard AM (1996) Mutation Res 371:29-46.

Diverse chemicals

Mean LED of GeneTox endpoints: SAL TA100, TA98, SCE, MLA, CHO



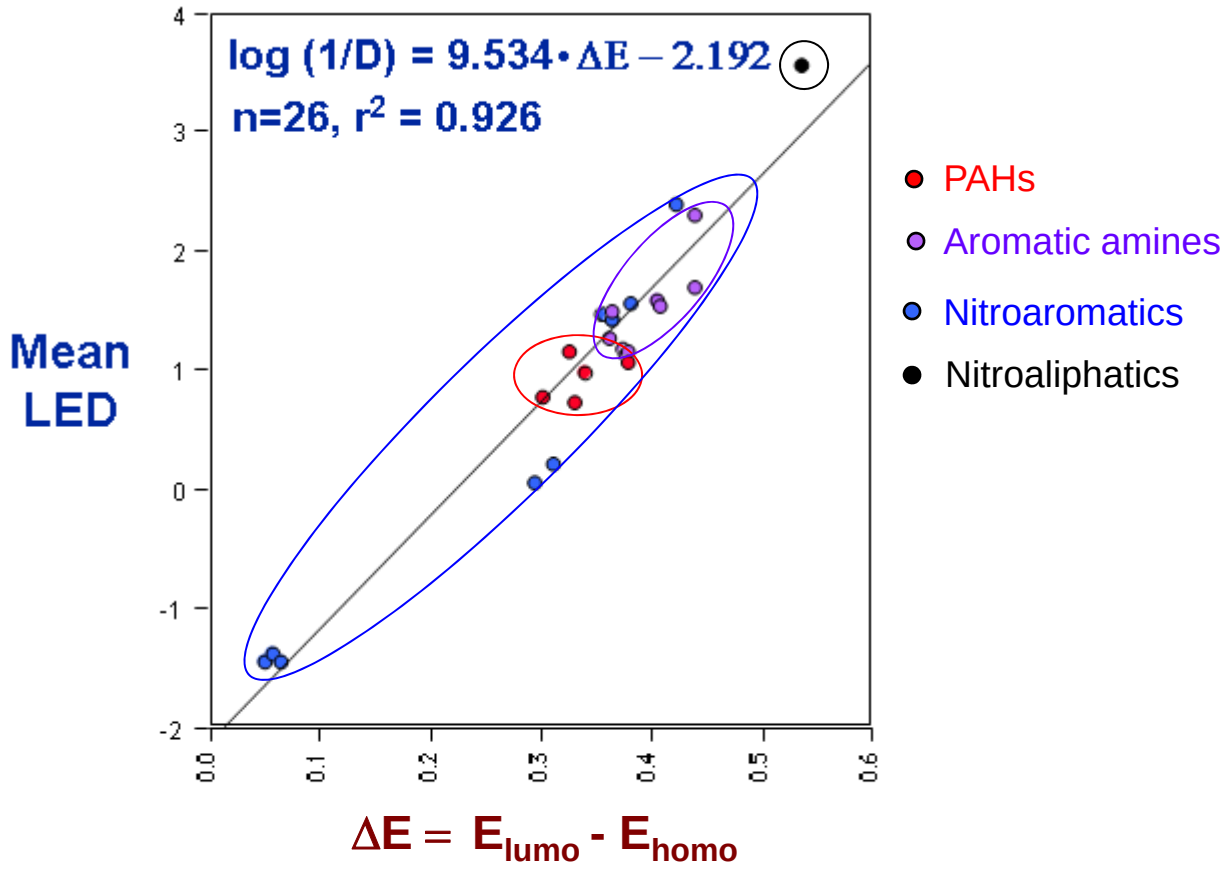
Question: *Can a QSAR predict gross mean potency across multiple genotoxicity measures and diverse chemical space?*

QSARs of mutagens & carcinogens: Case study illustrating problems in the construction of models for noncongeneric chemicals

Benigni R, Richard AM (1996) Mutation Res 371:29-46.

Diverse chemicals

Mean LED of GeneTox endpoints: SAL TA100, TA98, SCE, MLA, CHO



Question: *Can a QSAR predict mean potency across multiple genotoxicity assays and chemical classes?*

United States Environmental Protection Agency

QSARs of mutagens & carcinogens: Case study illustrating problems in the construction of models for noncongeneric chemicals

Benigni R, Richard AM (1996) Mutation Res 371:29-46.

Diverse chemicals

Mean LED of GeneTox endpoints: SAL TA100, TA98, SCE, MLA, CHO

Good statistics of initial QSAR primarily due to nitroaromatics in SAL TA100

Activity	NO_2	PAH	NH_2	n=26	n=60
SA0	0.93 0.85	0.43	0.21	0.89	0.49
SA9	0.93 0.94		0.30	0.93	0.62
GT5	0.89			0.85	0.0
SIC	0.77 0.70		0.16	0.56	0.11
Mean LED	0.95	0.38	0.49	0.93	

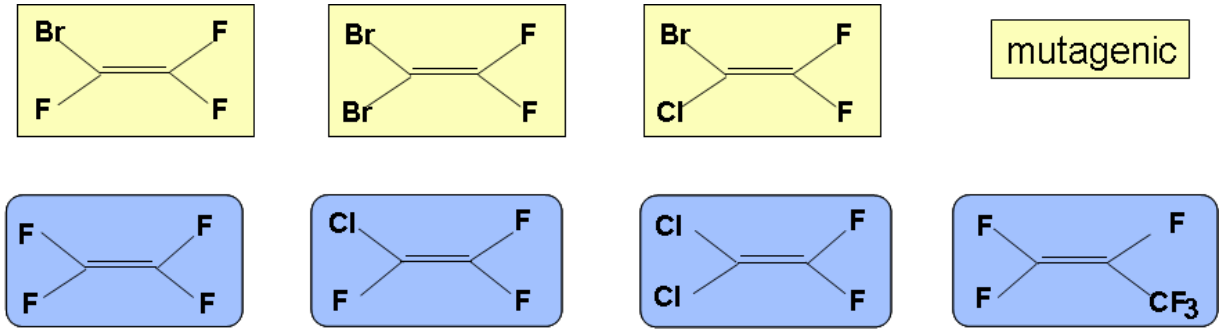
Question: *Can a QSAR predict mean potency across multiple genotoxicity assays and chemical classes?*

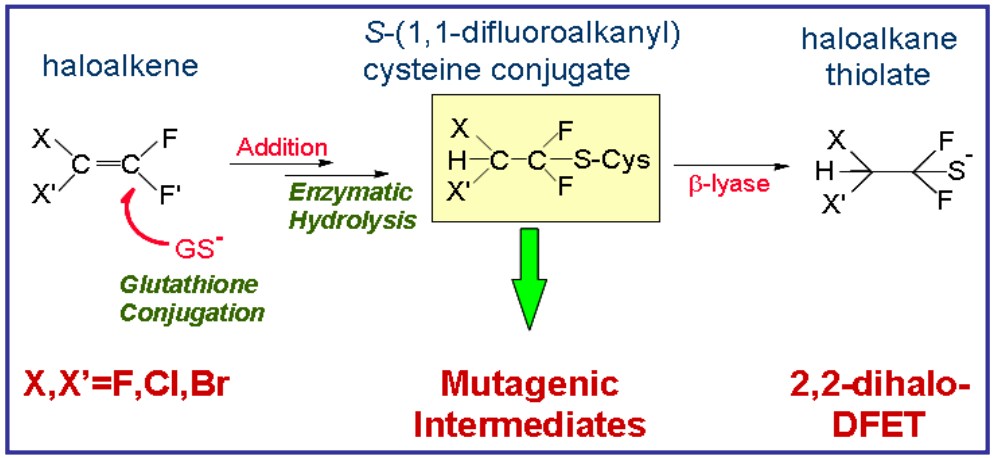
“Asking the right questions takes as much skill as giving the right answers.”

- Robert Half

Theoretical Evaluation of Two Plausible Routes for Bioactivation of S-(1,1-Difluoro-2,2-dihaloethyl)-L-cysteine Conjugates: Thiirane vs Thionoacyl Fluoride Pathway

Shim JY, Richard AM (1997) Chem Res Toxicol 10: 103-110.



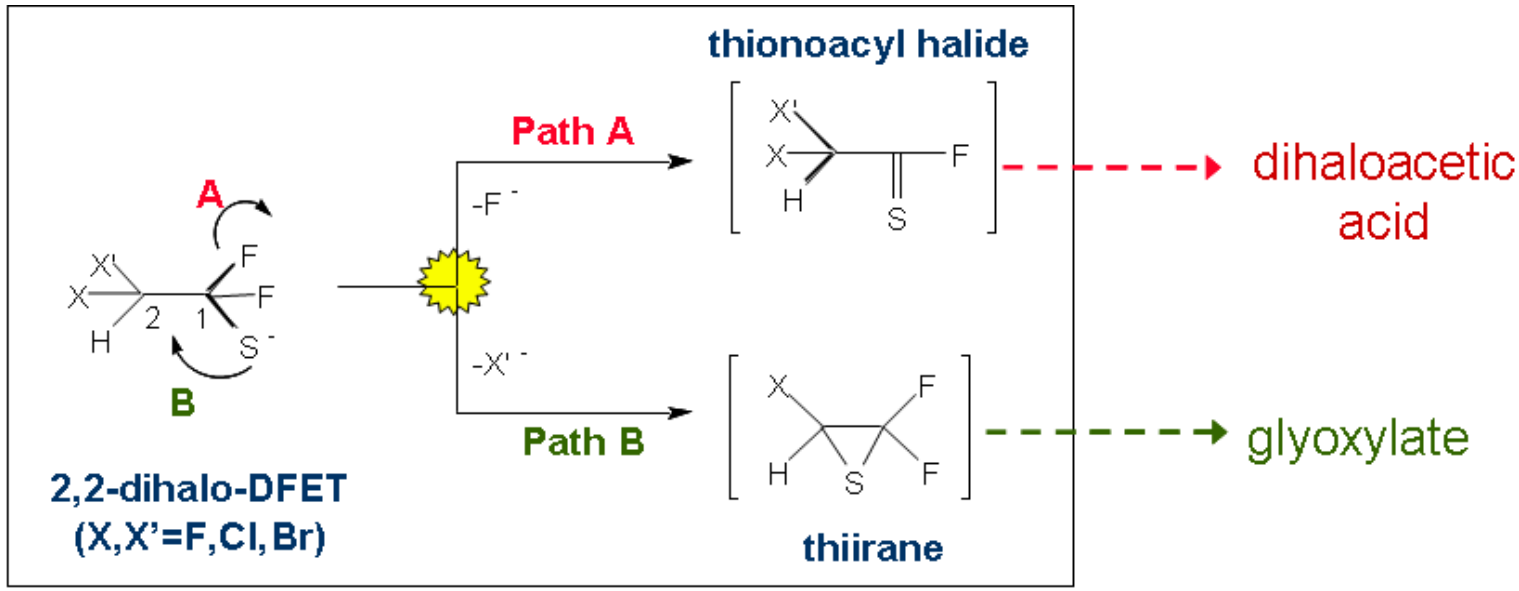


Given known glutathione activation mechanism, can different reactive intermediates account for the greater mutagenicity of brominated forms?

United States Environmental Protection Agency

Theoretical Evaluation of Two Plausible Routes for Bioactivation of S-(1,1-Difluoro-2,2-dihaloethyl)-L-cysteine Conjugates: Thiirane vs Thionoacyl Fluoride Pathway

Shim JY, Richard AM (1997) Chem Res Toxicol 10: 103-110.

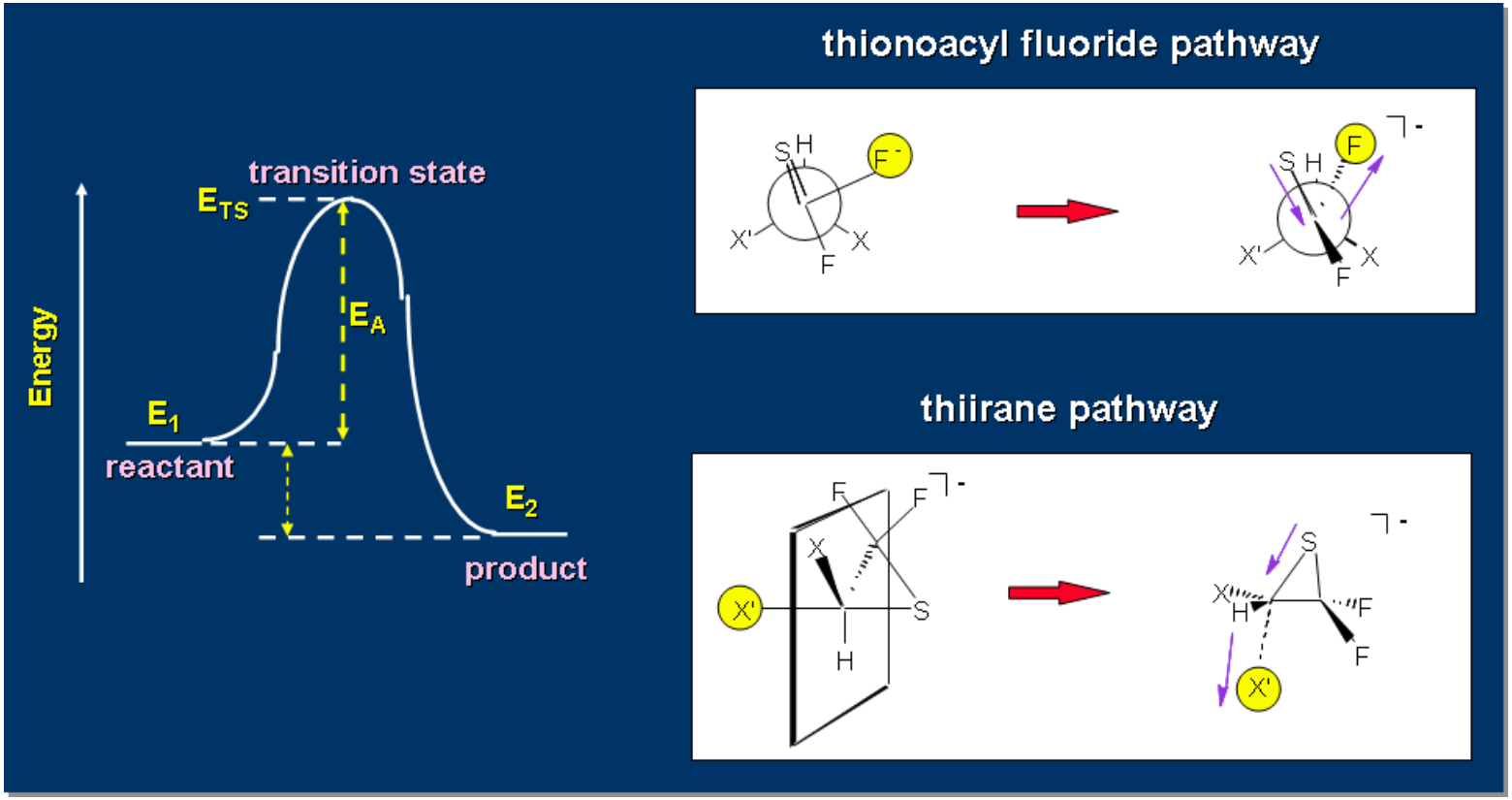


Given known glutathione activation mechanism, can different reactive intermediates account for the greater mutagenicity of brominated forms?

United States Environmental Protection Agency

Theoretical Evaluation of Two Plausible Routes for Bioactivation of S-(1,1-Difluoro-2,2-dihaloethyl)-L-cysteine Conjugates: Thiirane vs Thionoacyl Fluoride Pathway

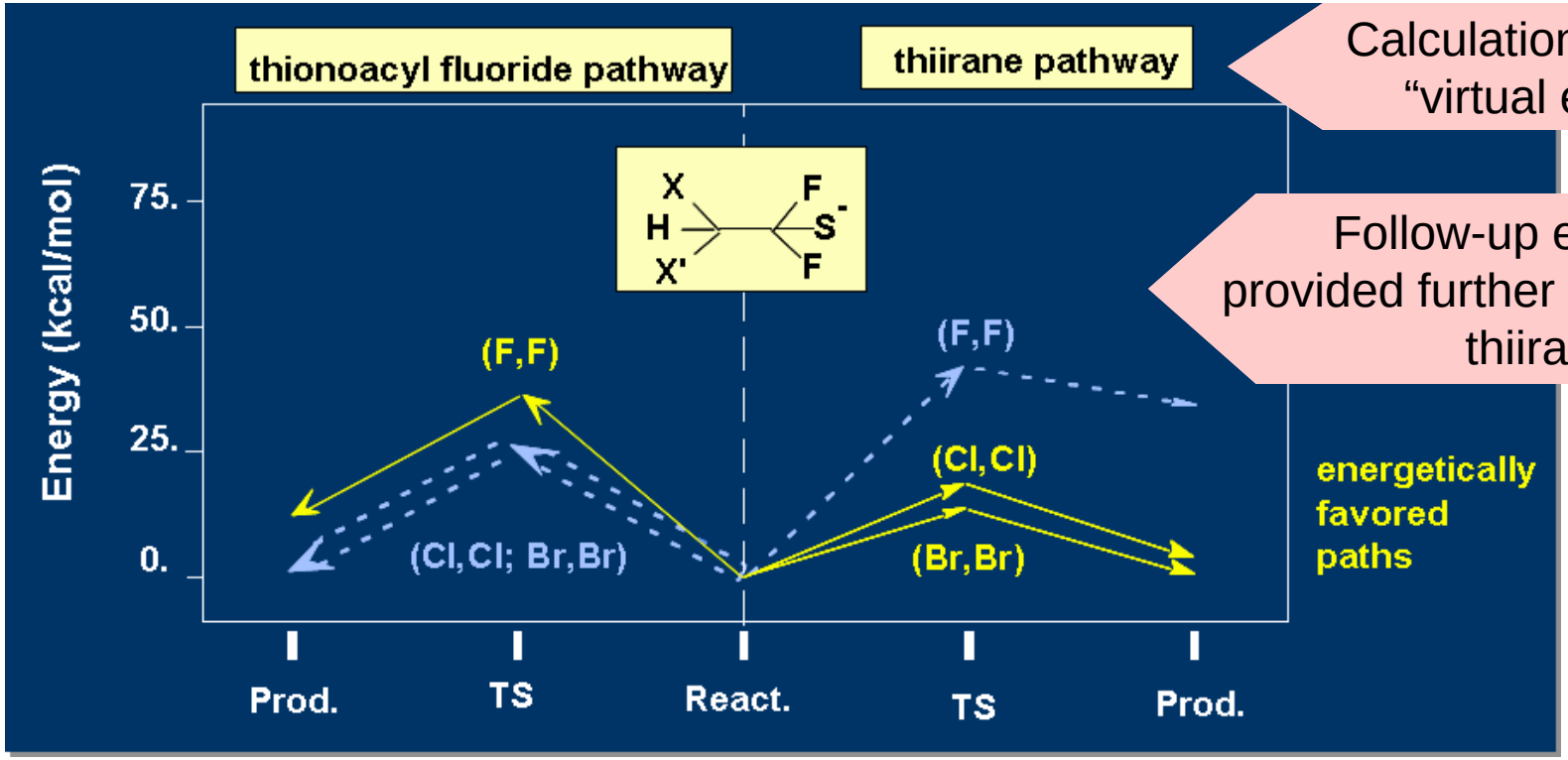
Shim JY, Richard AM (1997) Chem Res Toxicol 10: 103-110.



Given known glutathione activation mechanism, can different reactive intermediates account for the greater mutagenicity of brominated forms?

Theoretical Evaluation of Two Plausible Routes for Bioactivation of S-(1,1-Difluoro-2,2-dihaloethyl)-L-cysteine Conjugates: Thiirane vs Thionoacyl Fluoride Pathway

Shim JY, Richard AM (1997) Chem Res Toxicol 10: 103-110.



Calculation to perform "virtual experiment"

Follow-up experiments provided further evidence of thiirane pathway

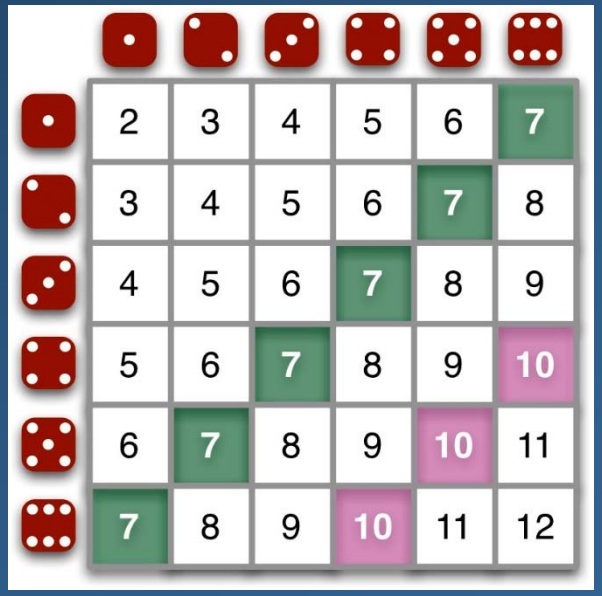
MP2/6-311+G**//HF/6-311G**

Given known glutathione activation mechanism, can different reactive intermediates account for the greater mutagenicity of brominated forms?

It all boils down to probabilities

"Important principles may and must be *flexible*."

- Abraham Lincoln



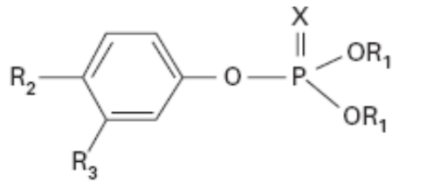
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.

agonist: makes enzyme work
antagonist: prevents enzyme working

Competitive AR antagonists



Chemical name[a]	No.	R_1	R_2	R_3	X	K_B (x 10^{-8} M)
Fenitrothion	1	CH_3	NO_2	CH_3	S	2.18[b]
	2	CH_3	NO_2	CH_3	O	ND[c]
Methylparathion	3	CH_3	NO_2	H	S	35.9
	4	CH_3	NO_2	H	O	ND
	5	C_2H_5	NO_2	CH_3	S	16.5
Ethylparathion	6	C_2H_5	NO_2	H	S	ND
	7	C_2H_5	CH_3	H	S	ND
	8	C_2H_5	H	CH_3	S	ND
	9	C_2H_5	$(O–CH_2–O)$		S	ND
	10	n-C_3H_7	NO_2	CH_3	S	ND
Flutamide[d]						1.07[e]
Hydroxyflutamide[d]						0.22
Linuron[a]						75.8[f]

Known antagonists

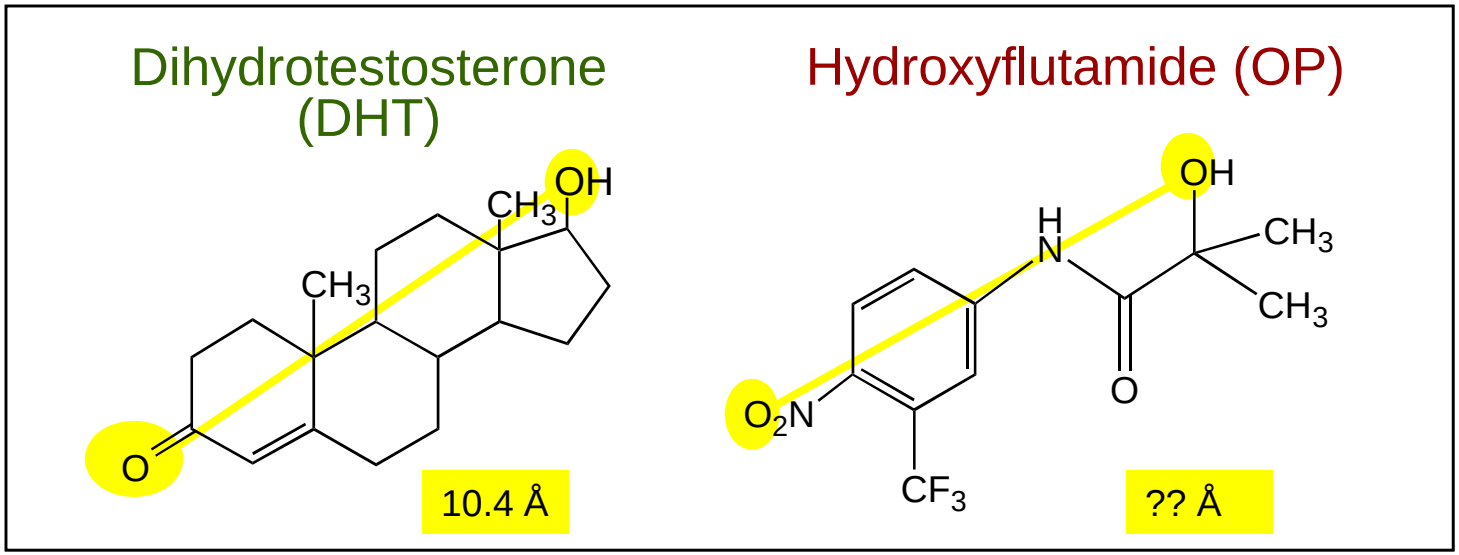
Fenitrothion & Hydroxyflutamide are both weak agonists in the absence of DHT

Fenitrothion and Flutamide have similar antagonist potency

What distinguishes the AR antagonists-only from the antagonists/agonists within a series of organophosphates?

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.



2D ⇒ 3D:

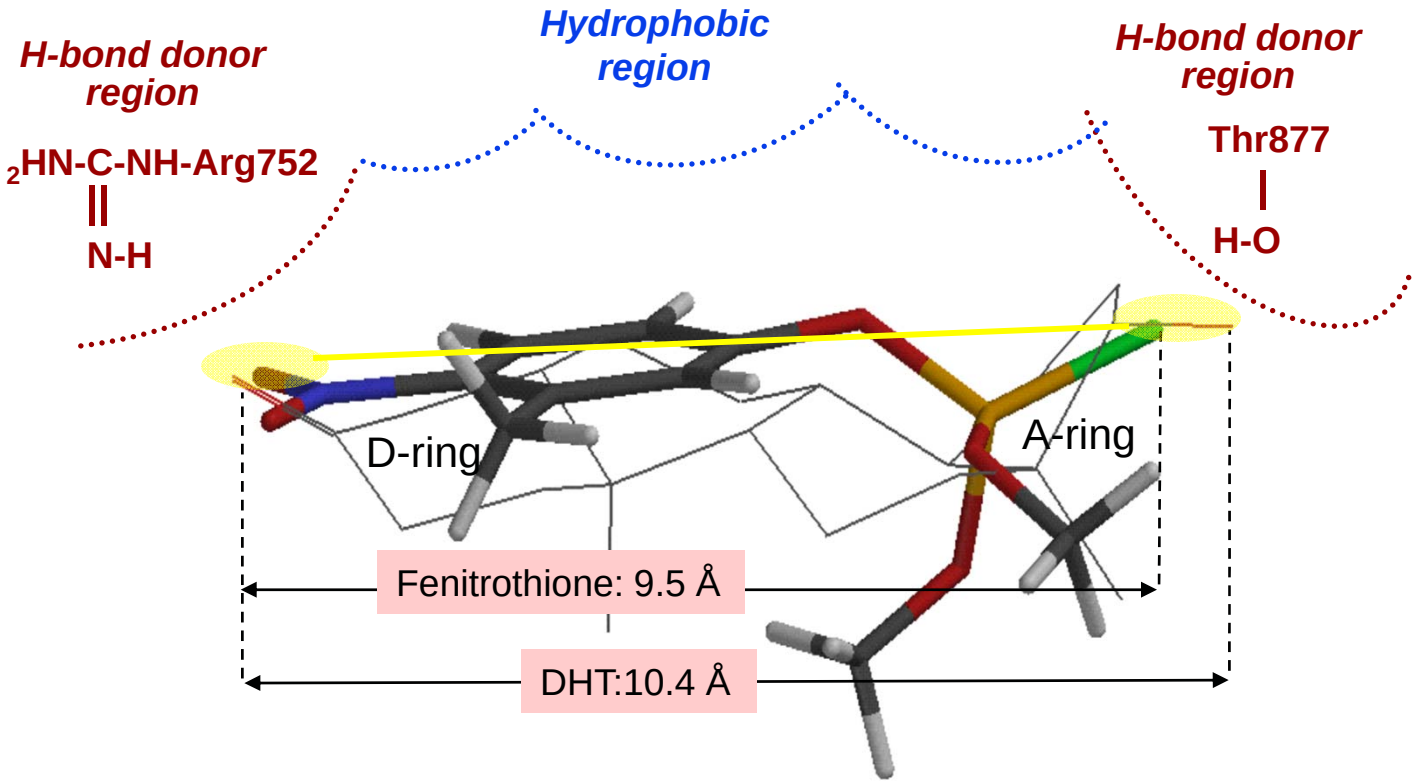
- Compute max distance between polar groups for OPs
- Compute ΔE to reach stretched conformation
- Compute H-bonding interaction energies & distances

Translate question to computational problem

What distinguishes the AR antagonists-only from the antagonists/agonists within a series of organophosphates?

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.



Is it feasible (i.e., low energy cost) for the OP to achieve the H-bonding configuration of DHT?

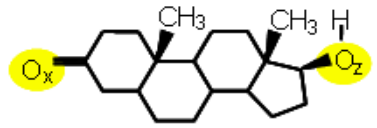
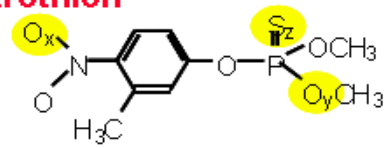
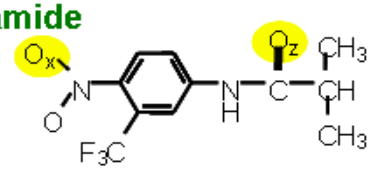
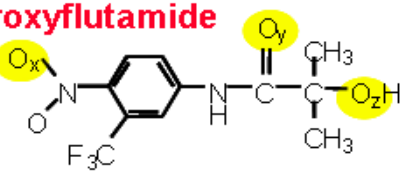
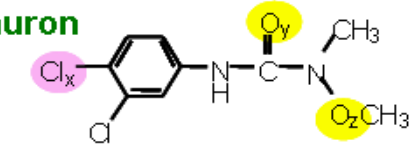
United States Environmental Protection Agency

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.

Agonists can achieve DHT binding distance and effective H-bonds

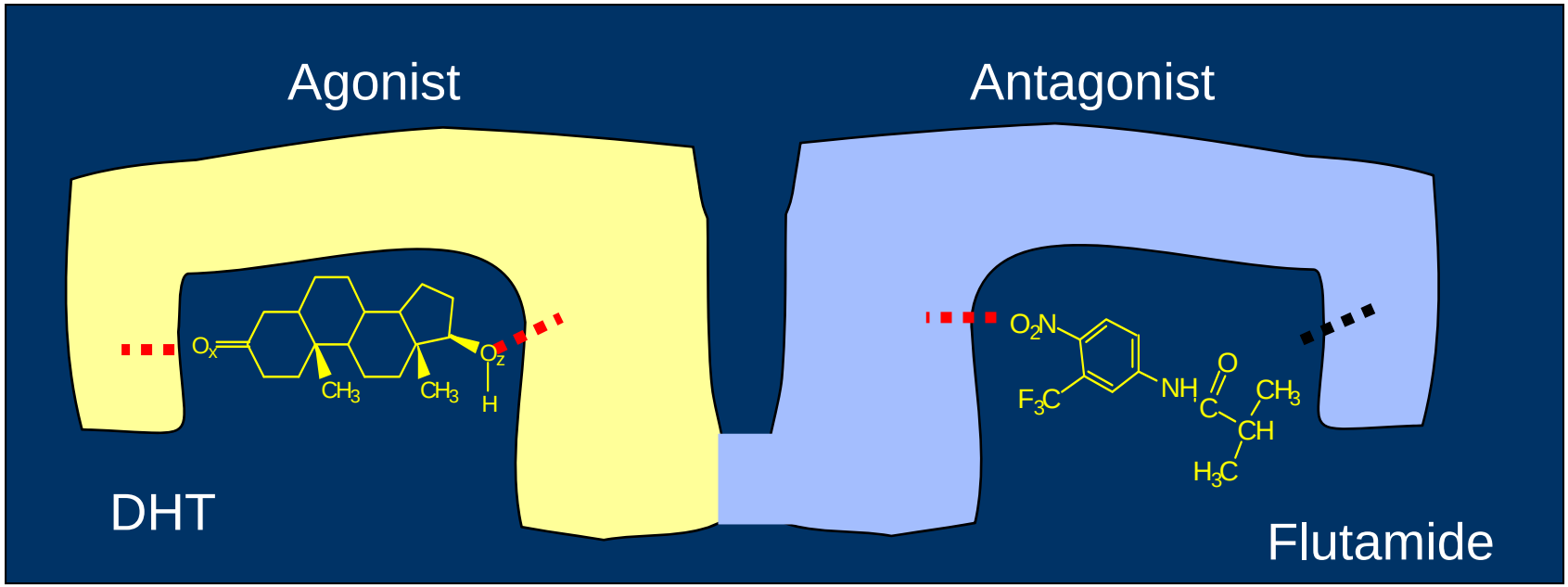
Antagonists can't

Chemical Structure Agonist/Antagonist	Energy (kcal/mol)	Distance(Å) x - - - z	H-bond (kcal/mol) x	y	z
DHT 	0.00 0.36	10.48 10.00	-3.70 (1.81Å)		-3.14 (1.82Å)
Fenitrothion 	0.00 1.68	8.18 9.50 (8.09)	-3.33 (1.83Å)	-0.17 (1.88Å)	-2.17 (2.83Å)
Flutamide 	0.00 2.48	7.50 8.00	-2.91 (1.85Å)		-2.97 (1.82Å)
Hydroxyflutamide 	0.00 1.72	8.94 9.50 (7.27)	-2.96 (1.83Å)	-3.20 (1.81Å)	-2.95 (1.83Å)
Linuron 	0.00 1.65	9.16 9.50 (7.00)	-2.06 (2.64Å)	-3.25 (1.82Å)	-2.85 (2.68Å)

Is it feasible (i.e., low energy cost) for the OP to achieve the H-bonding configuration of DHT?

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.



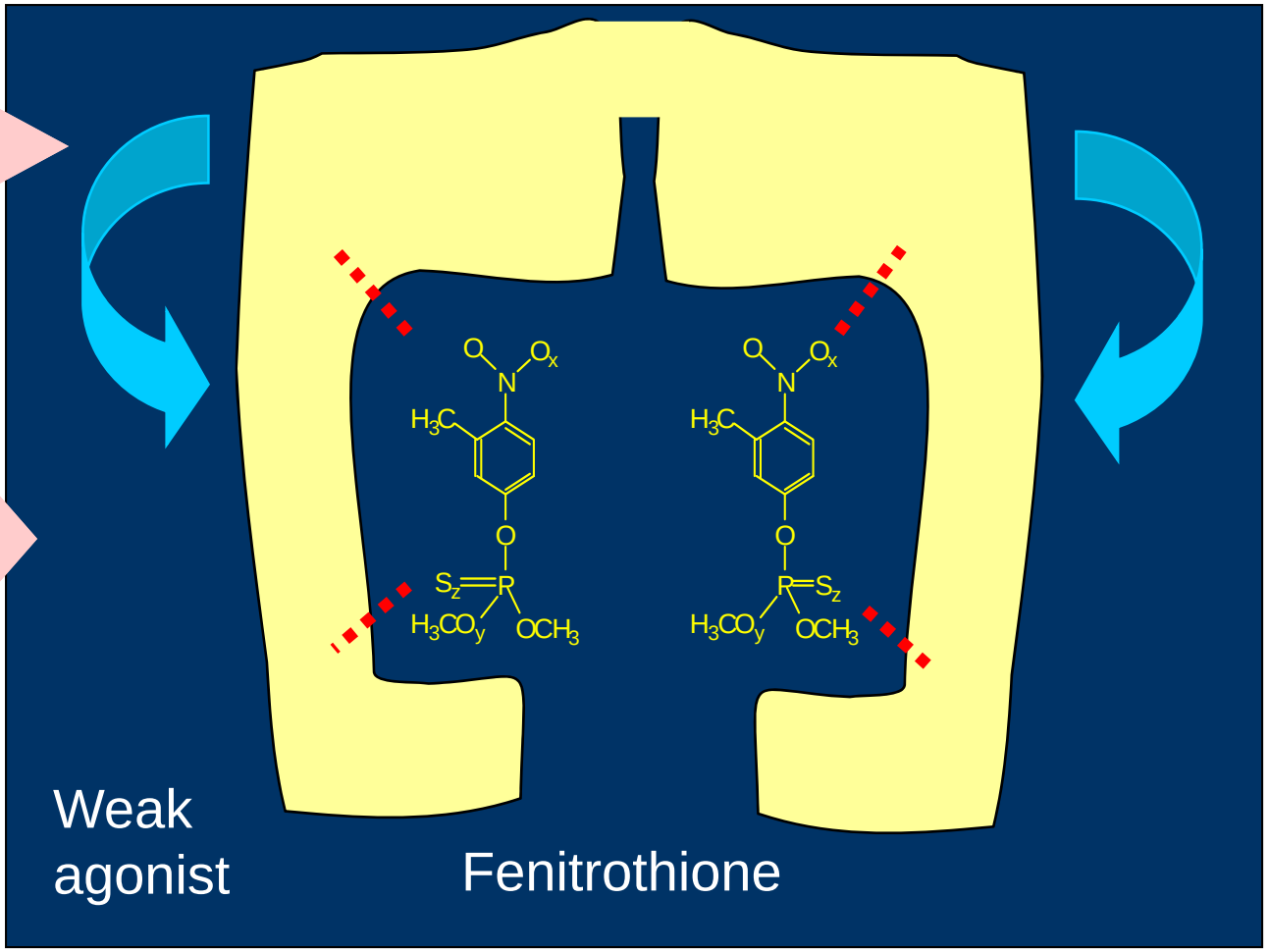
Low dose α
High probability event

Interaction of Organophosphate Pesticides and Related Compounds with the Androgen Receptor

Tamura H, Yoshikawa H, Gaido KW, Ross SM, DeLisle RK, Welsh WJ, Richard AM (2003) EHP 111:1-8.

High dose α
Low probability event

Relatively unsophisticated calculations can lend insight into target interaction



Lessons learned ...

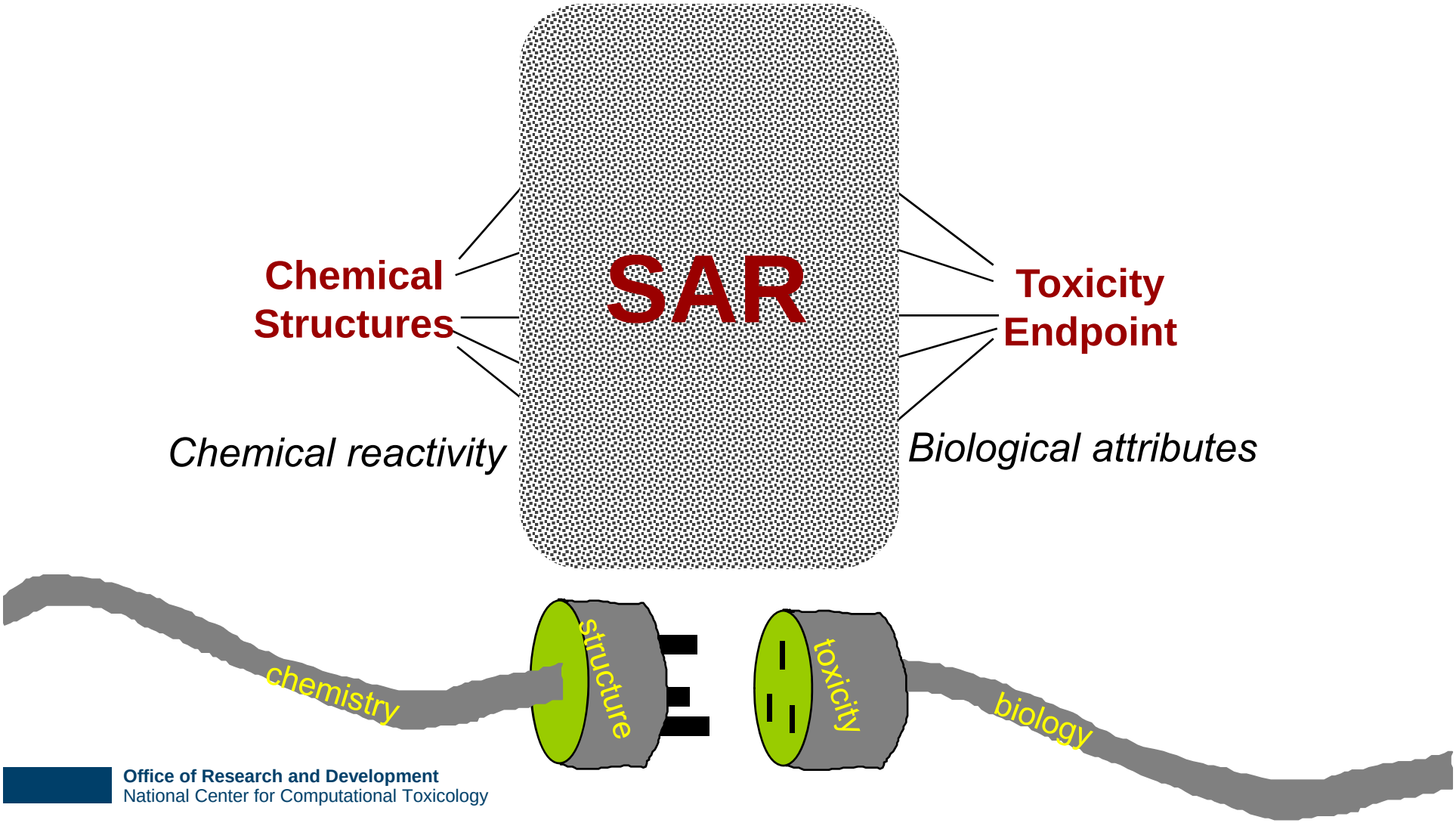
- Important for modelers to collaborate with “domain” experts (e.g., toxicologists)
- Computational chemistry & QSAR approaches can be effective partners to experiment
- SAR works best with similar chemicals and clear mechanisms
- Importance of asking the right questions, and using the appropriate level of theory for the problem

Toxicity prediction is not easy!

Part II:
Computational Toxicology

Data challenges &
cheminformatics approaches
to predicting toxicity

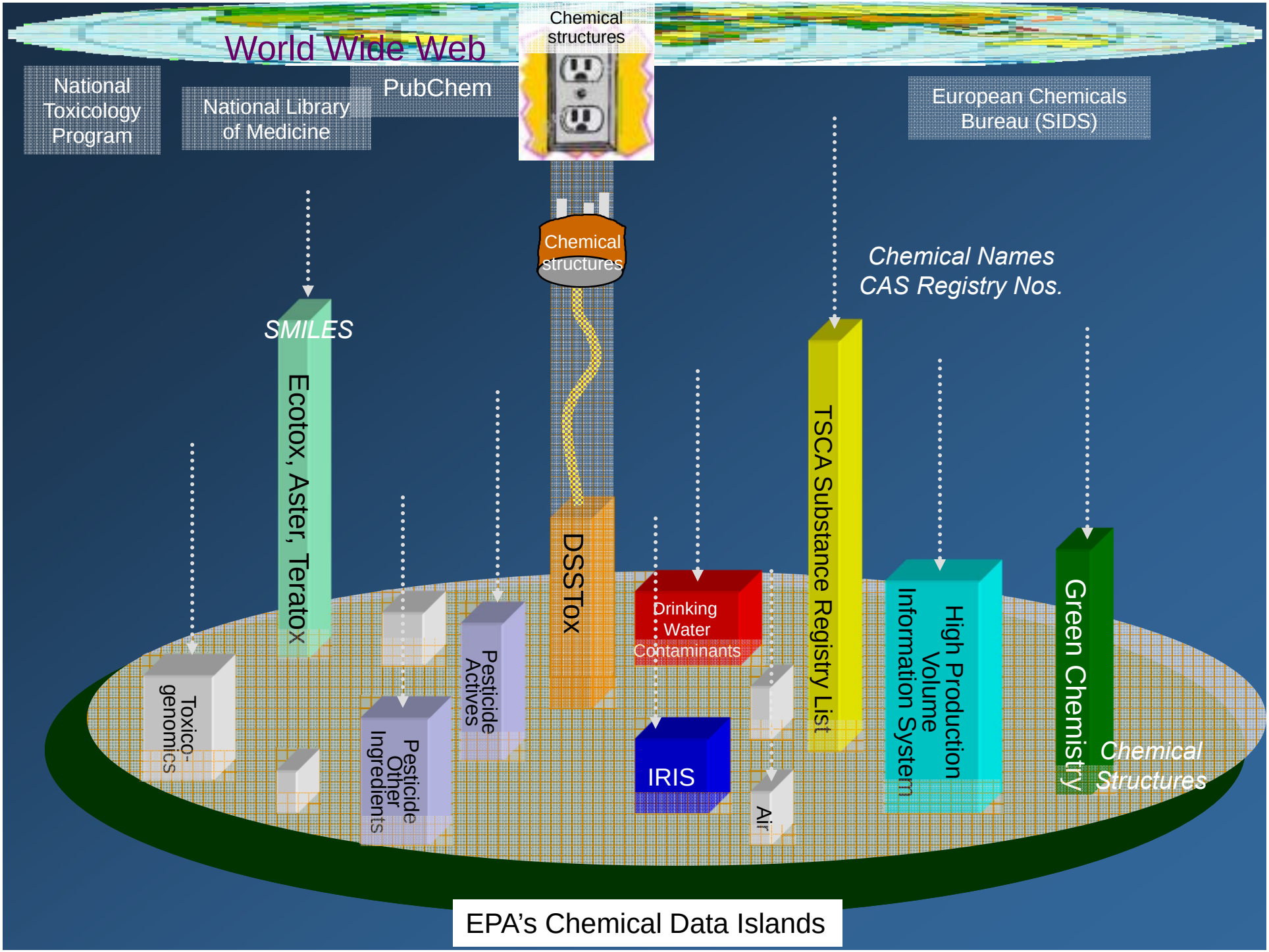
Global vs. Local SAR models



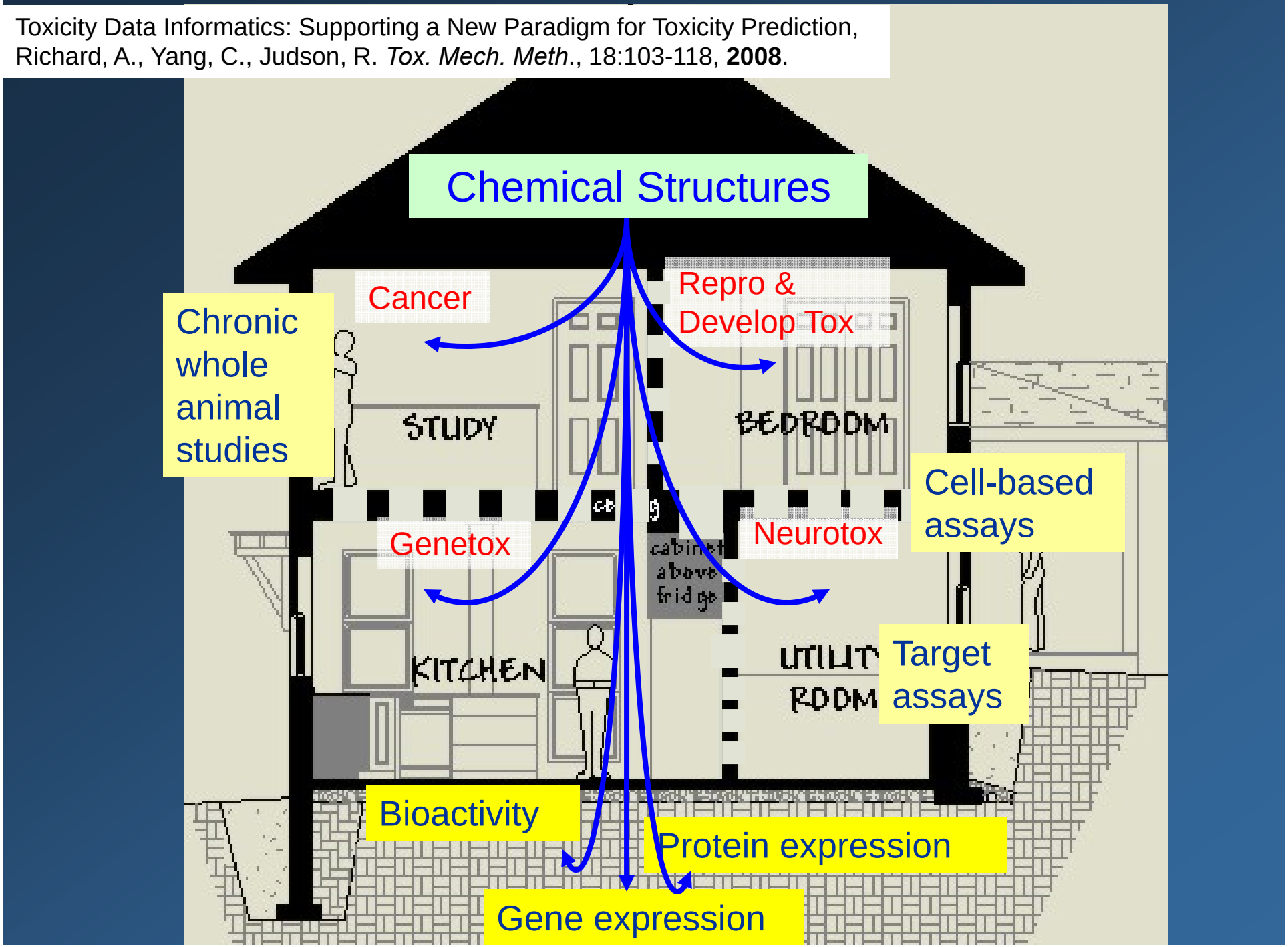
The bigger challenge -
predict potential toxicity of any chemical

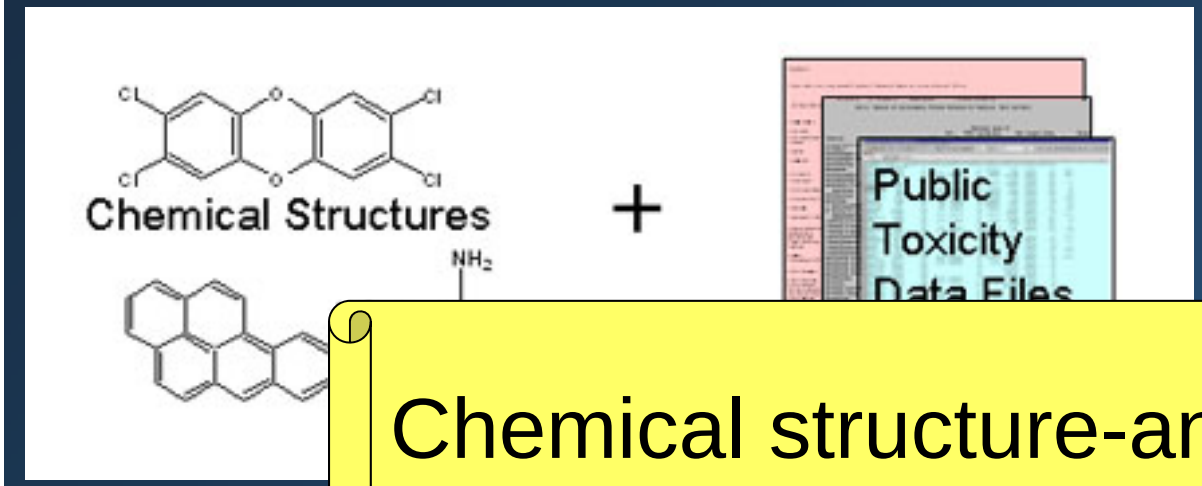
What are the chemicals of concern?
Where are the data to develop models??

EPA's Chemical Data Islands


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

Toxicity Data Informatics: Supporting a New Paradigm for Toxicity Prediction, Richard, A., Yang, C., Judson, R. *Tox. Mech. Meth*., 18:103-118, **2008**.





Chemical structure-annotation

Distribut**d**

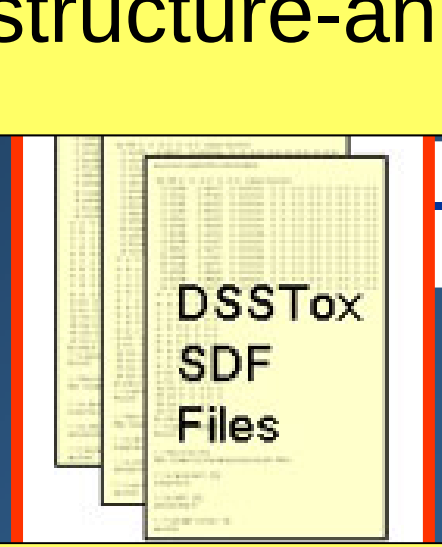
Structure-**S**earchable

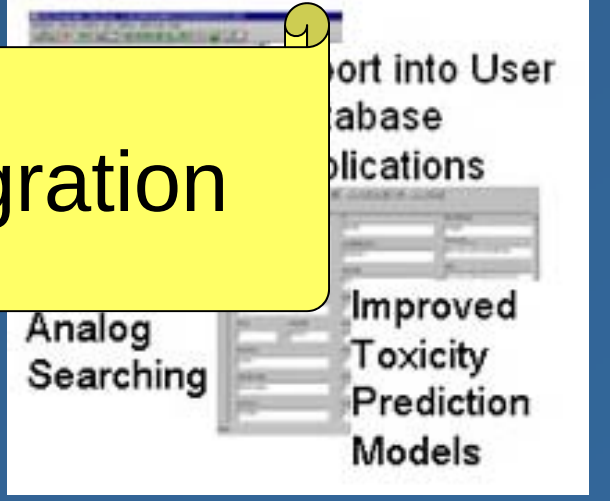
Toxicity

Public

Databas**e**

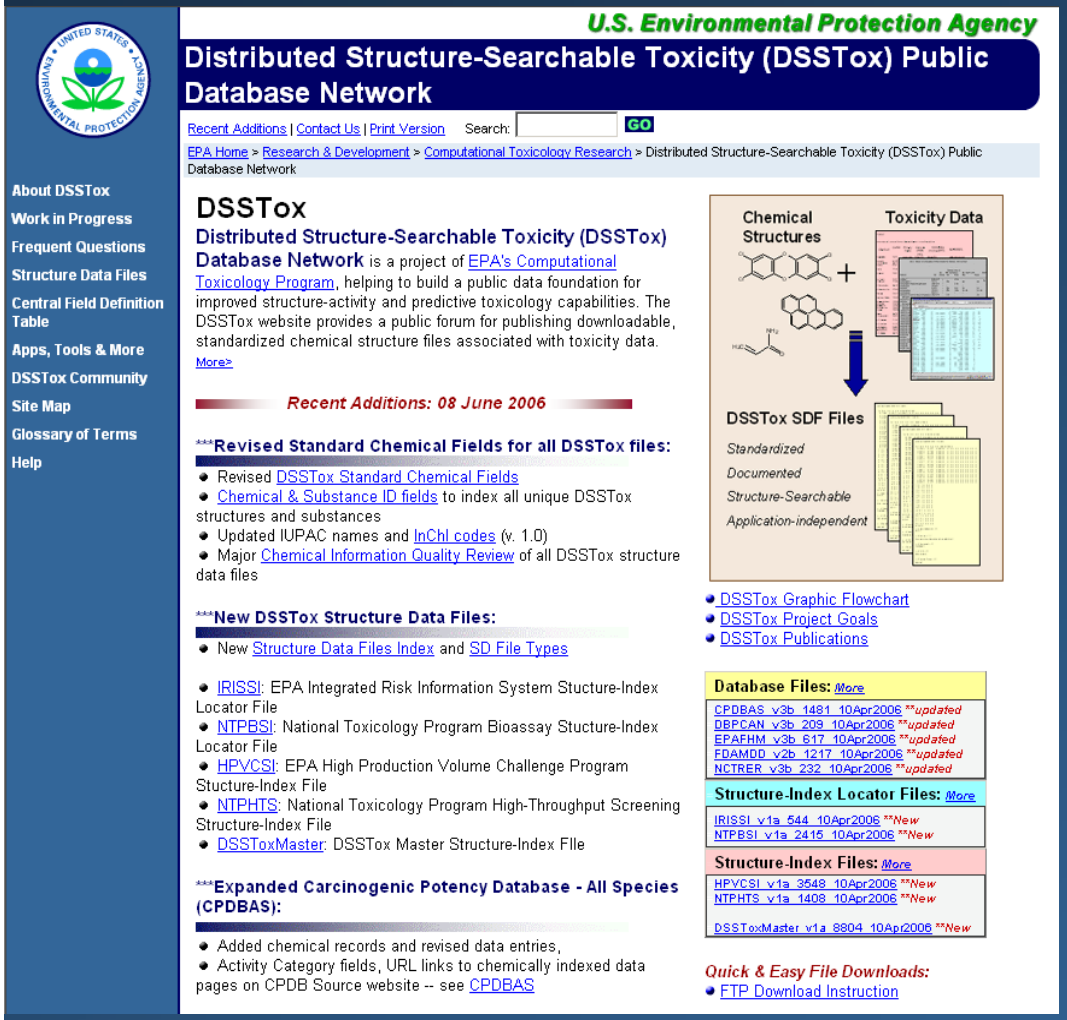
Network





Data standards and integration

EPA's DSSTox Public Website – Launched 2004



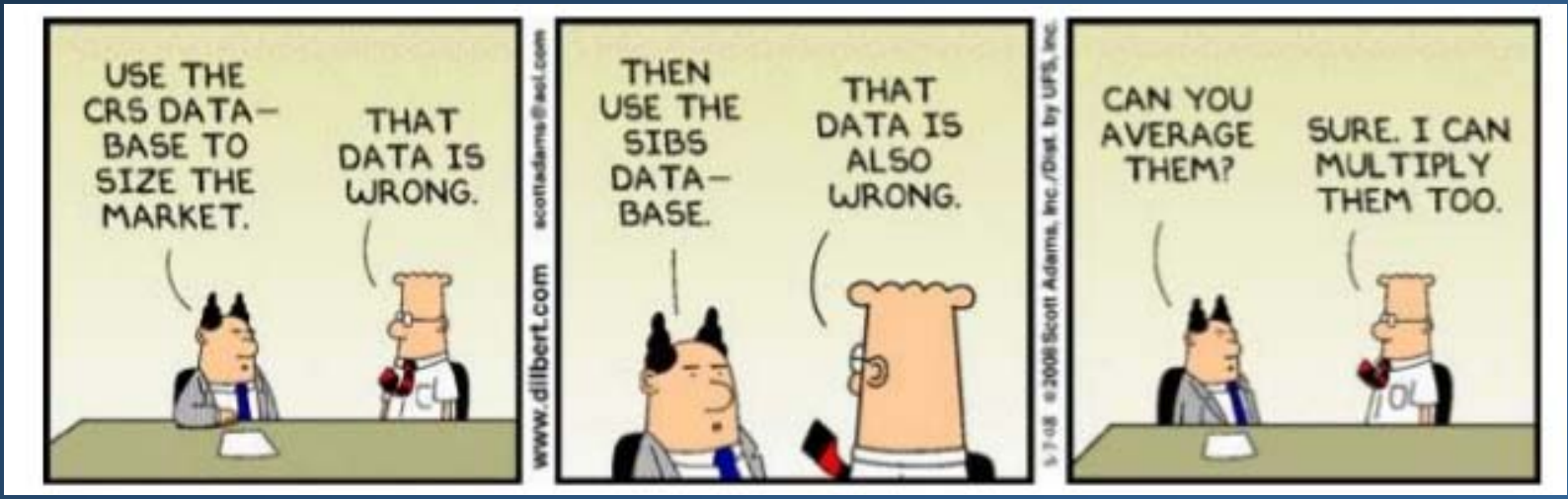
- Focus on environmental chemicals & EPA lists
- Toxicity datasets for building SAR models
- High quality, manually curated data-structure associations
- Downloadable structure files

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

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

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

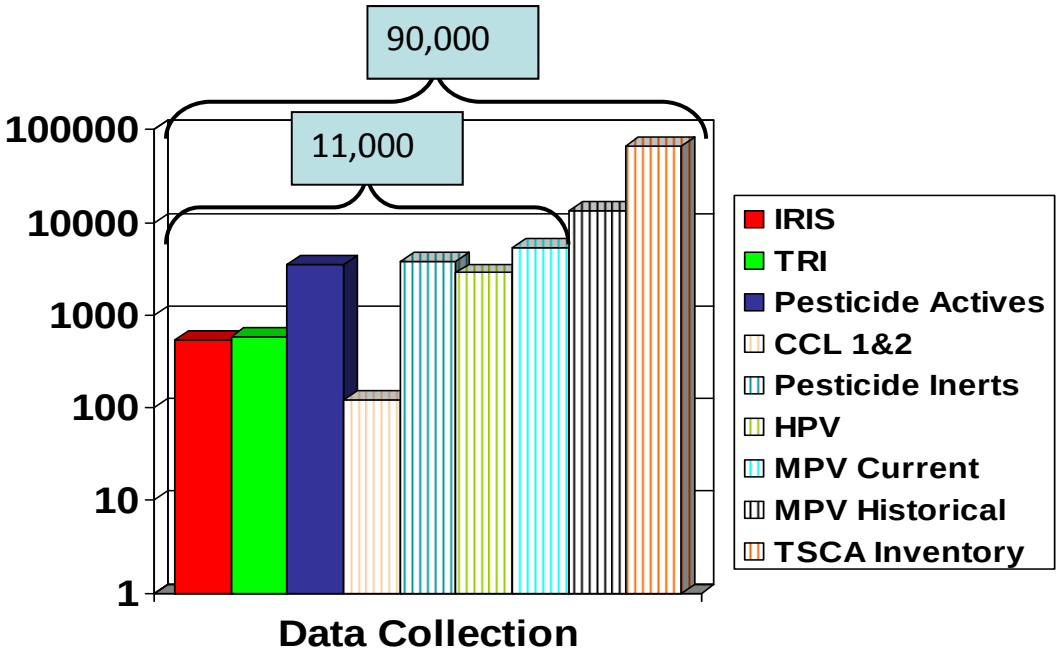
A prediction model is only as good as the data that goes into it …



Data quantity,
Data quality!

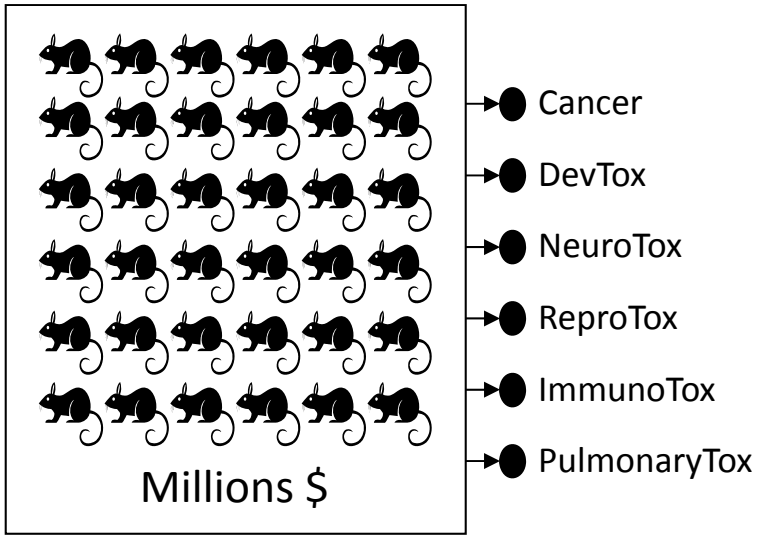
The Big Problem …..

Too Many Chemicals



…and not enough data.

Too High a Cost

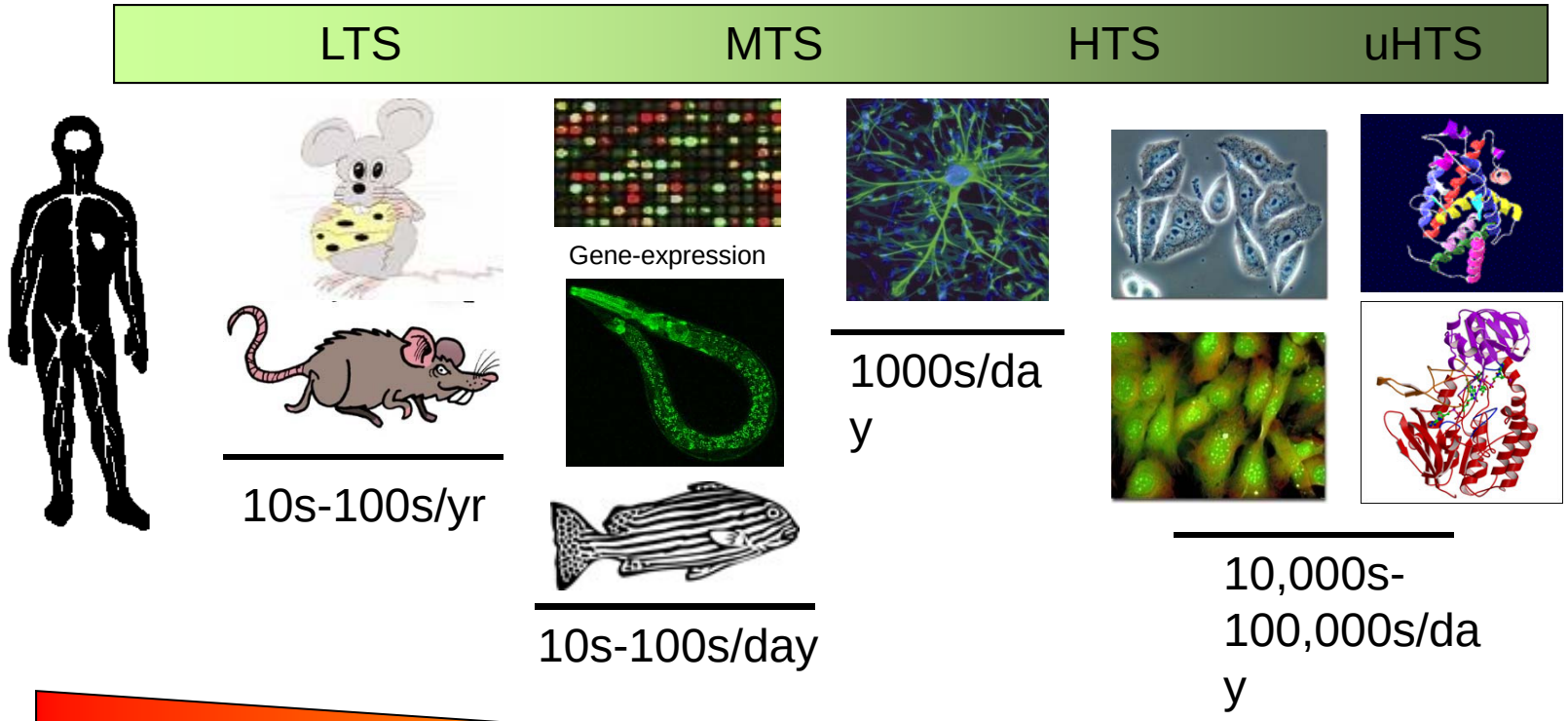


Too many endpoints
Too many mechanisms

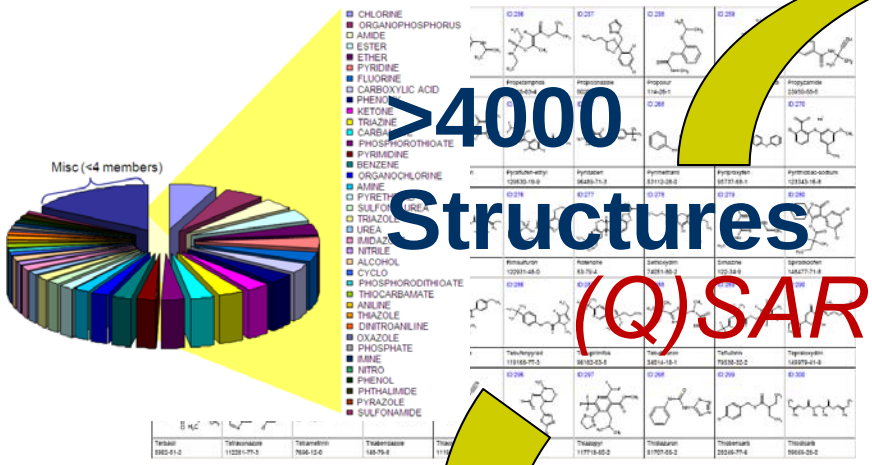
Judson, et al EHP, 2008

High-Throughput Screening Assays

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

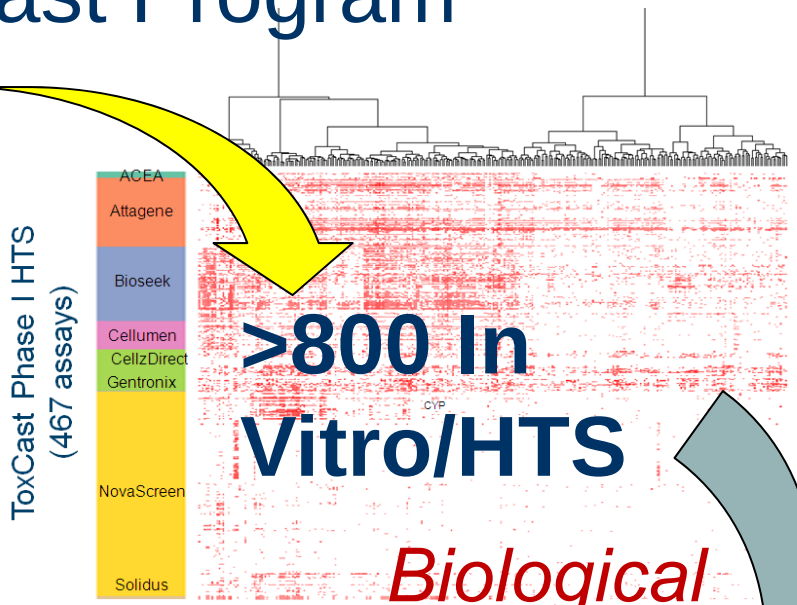


EPA's ToxCast Program



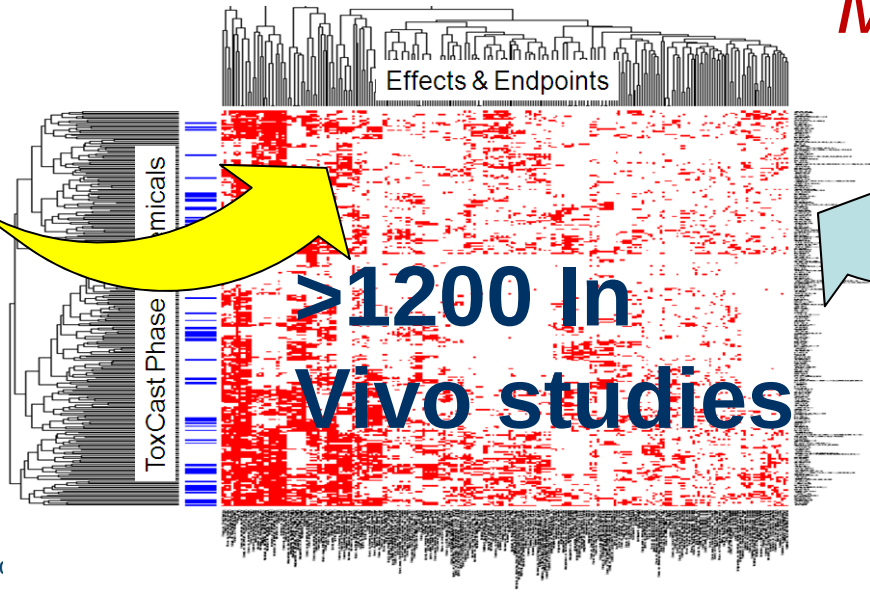
>4000 Structures

(Q)SAR



>800 In Vitro/HTS

Biological Modeling

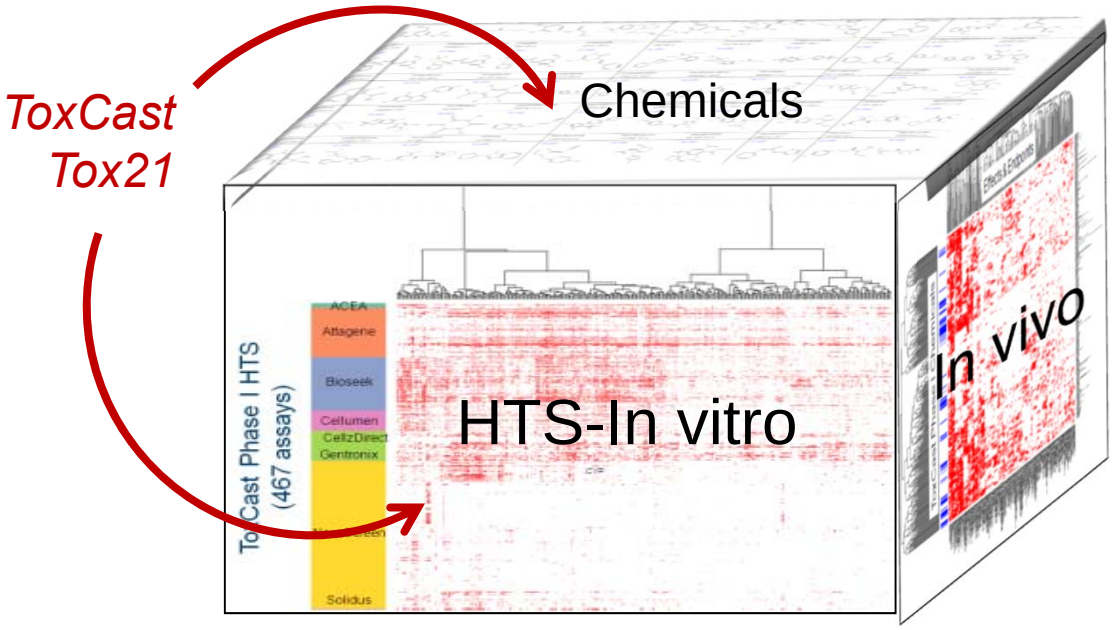


>1200 In Vivo studies

Lots of data!!!!

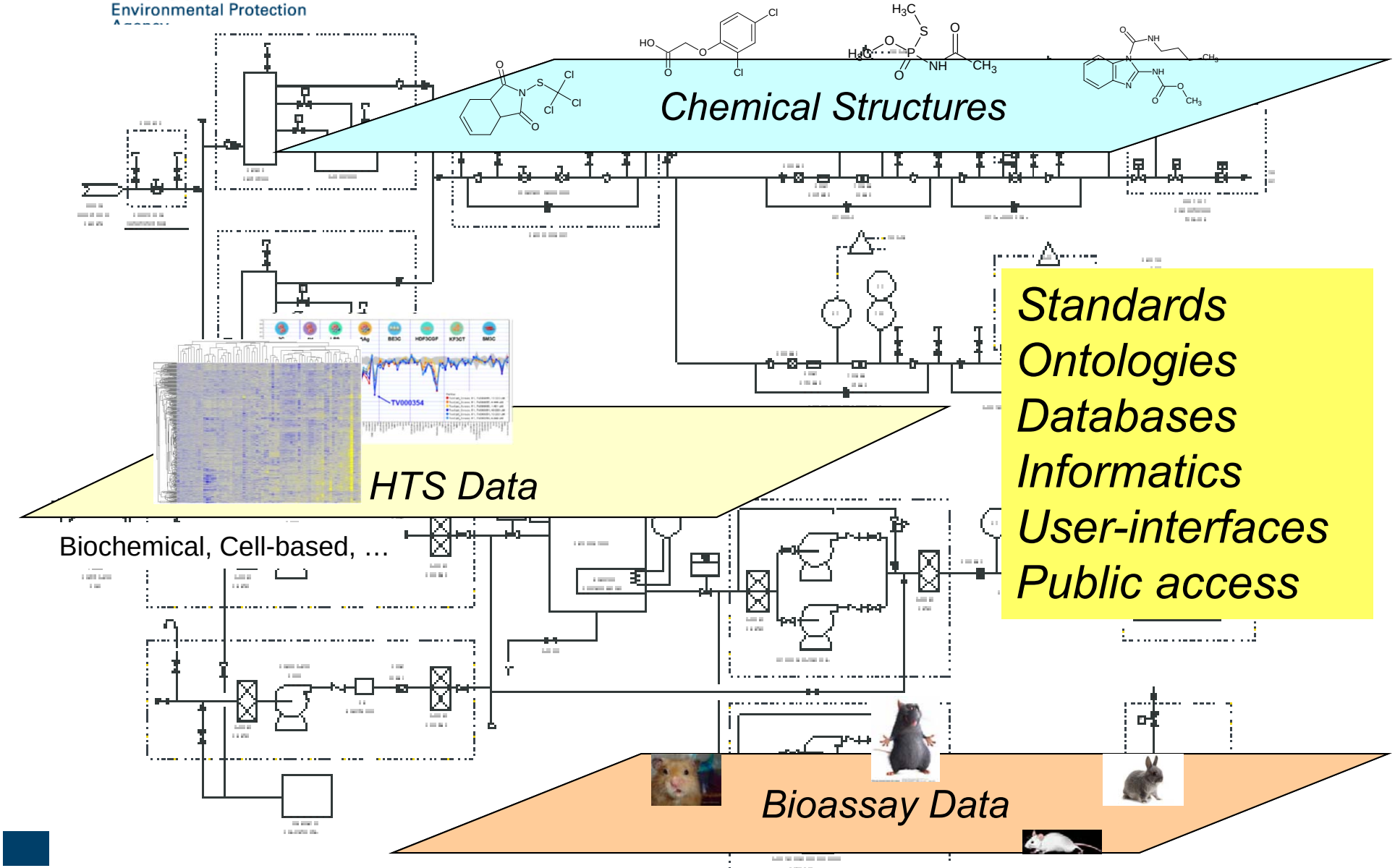
Chemical “probes” of biological activity

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



ToxCast: "Big Data" Informatics Challenges

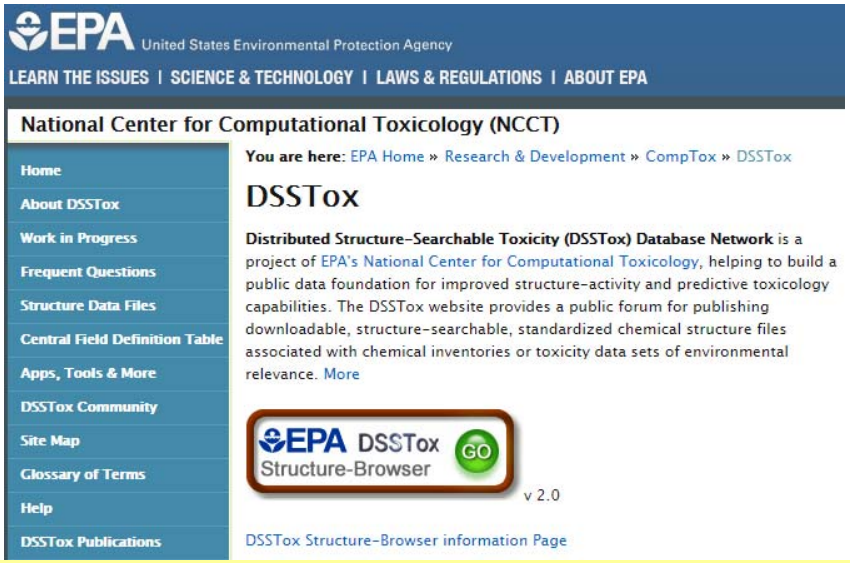
United States Environmental Protection Agency



Standards
Ontologies
Databases
Informatics
User-interfaces
Public access

DSSTox Update

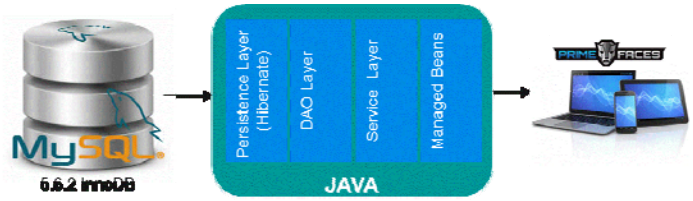
DSSTox_v1

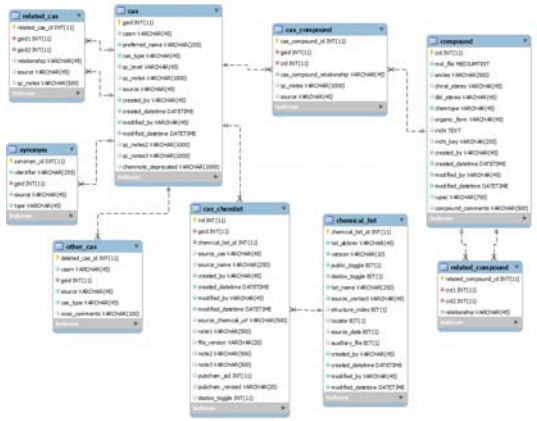


- Manually curated 25K substance records
- EPA-focus, environmental tox datasets
- Emphasis on accurate CAS-name-structure annotations
- Public resource for high-quality structure-data files (SDF)

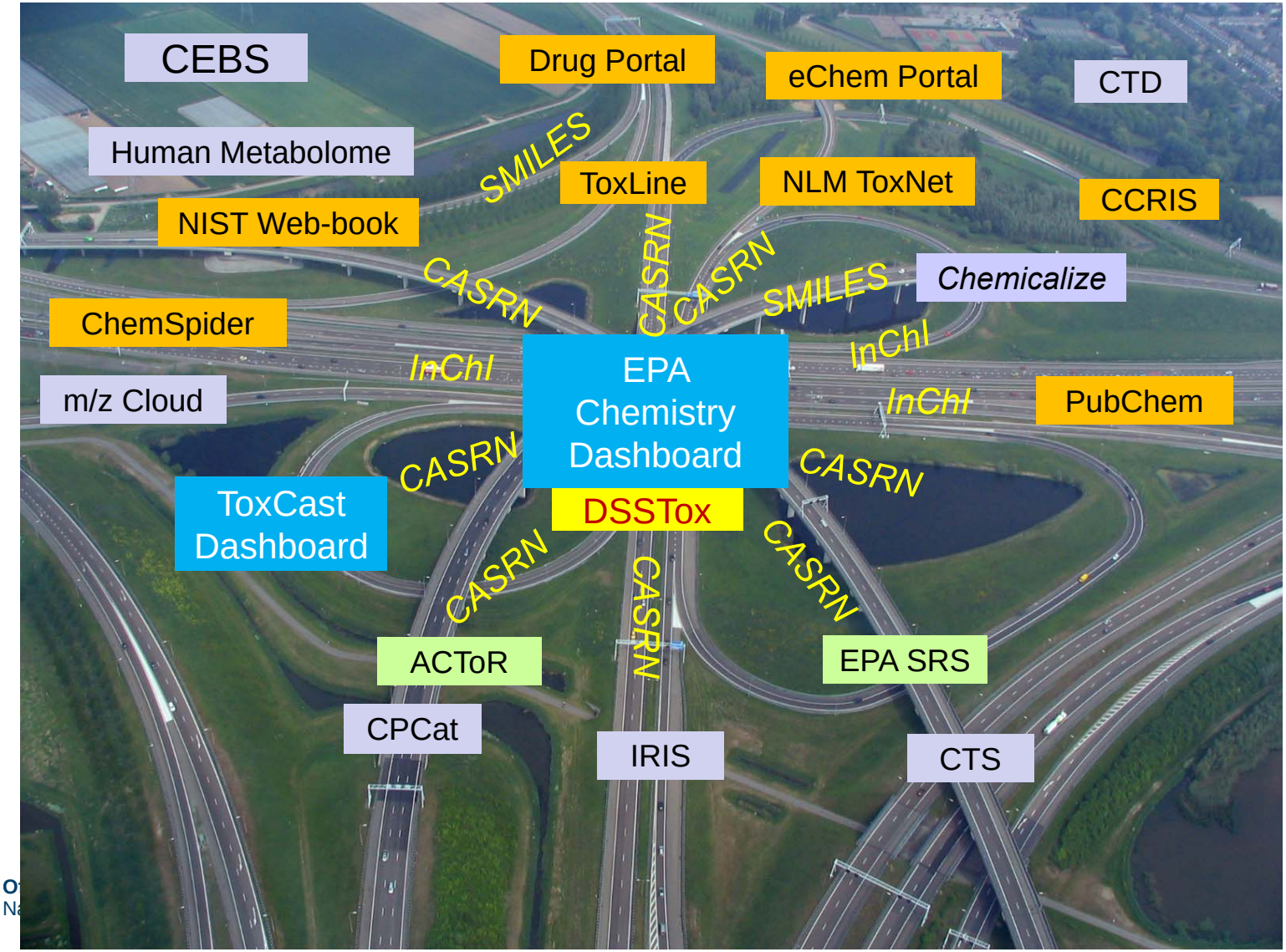
DSSTox_v2

- Convert DSSTox tables to MySQL
- Develop curation interface & cheminformatics workflow
- Expand chemical content to 750K
- Web-services & Dashboard access

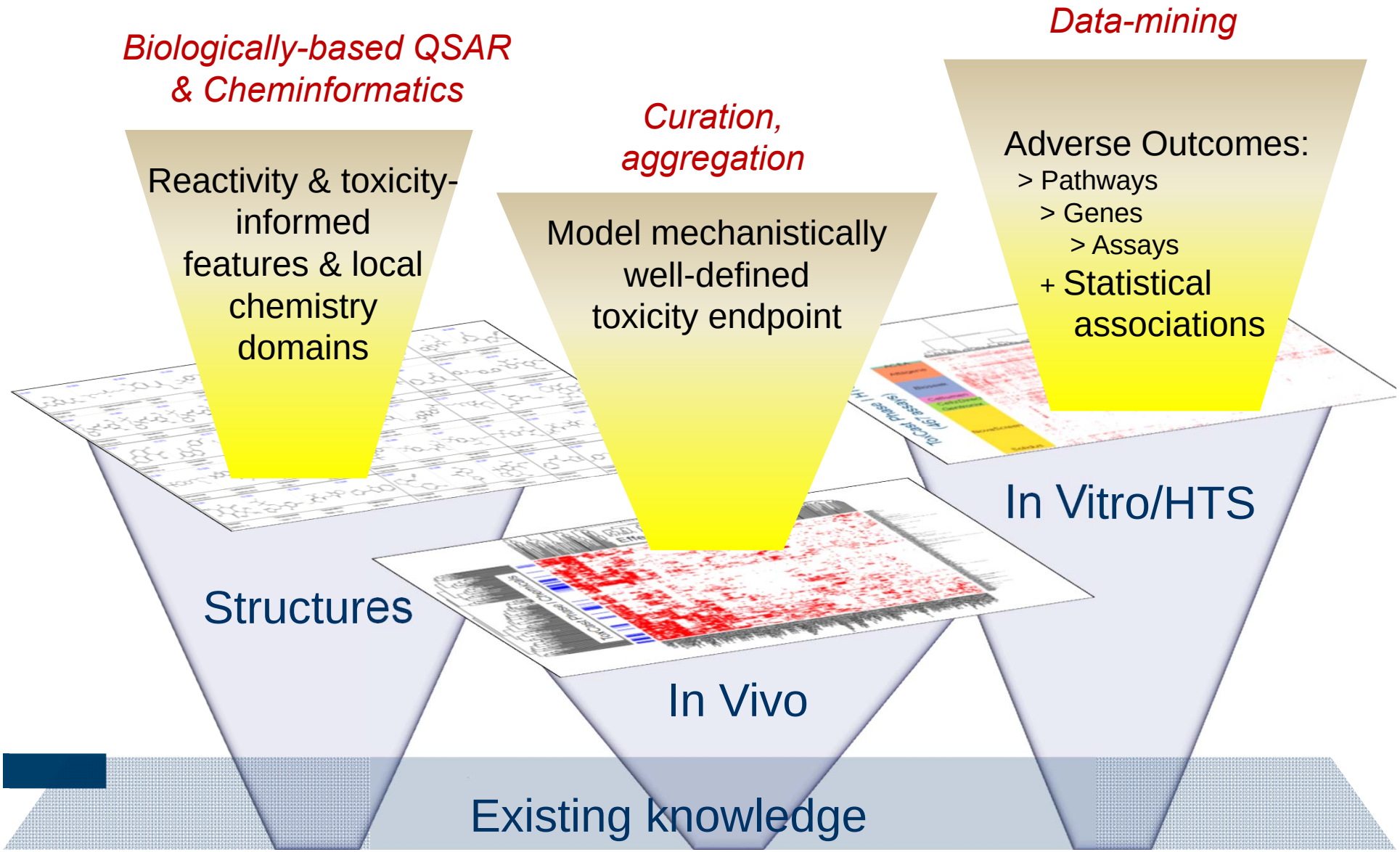




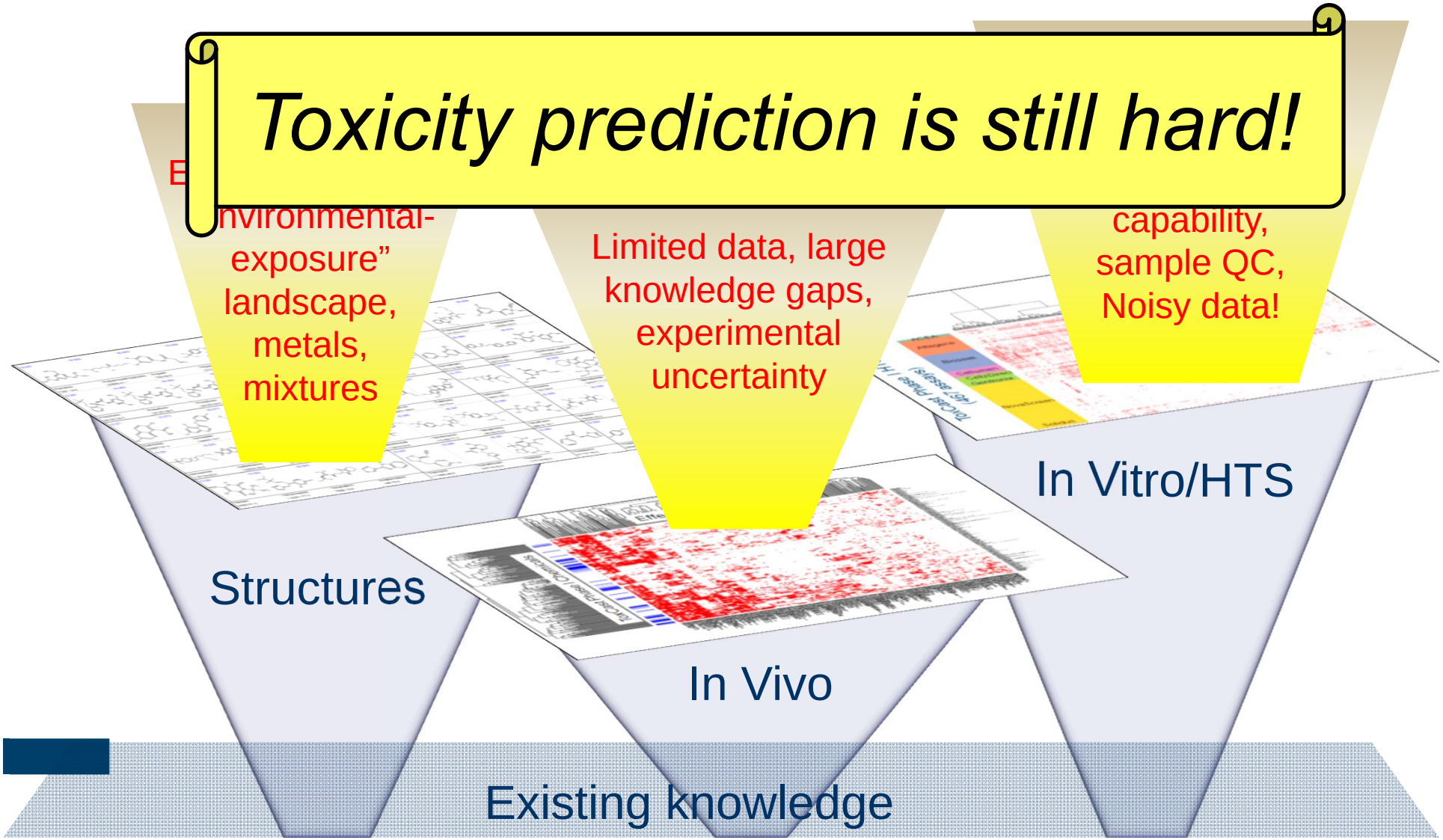
DSSTox is now EPA's chemical "hub"



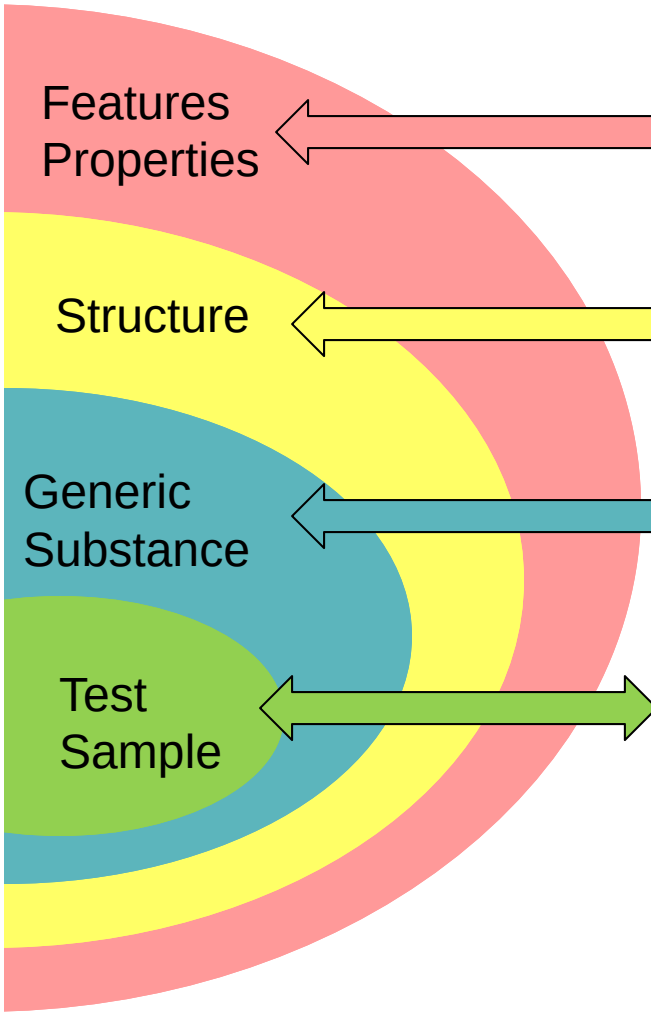
Toxicity Prediction Challenge



Toxicity Prediction Challenge

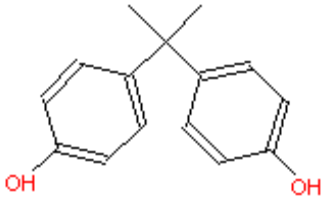


Chemical Representations supporting Data Integration



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

Chemical analogs, Read-across, SAR modeling



SMILES
InChI

Structure searching & modeling

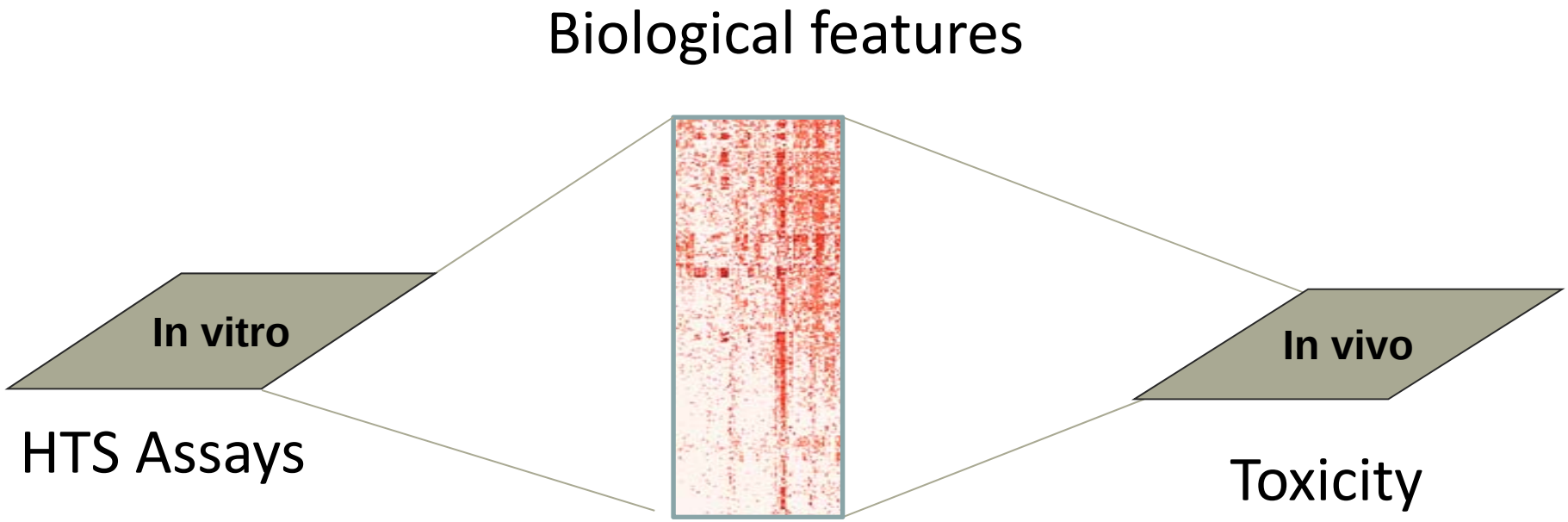
Chemical Name
CASRN

Public toxicity datasets

Supplier, Lot/Batch, physical description

Experimental Endpoint Data

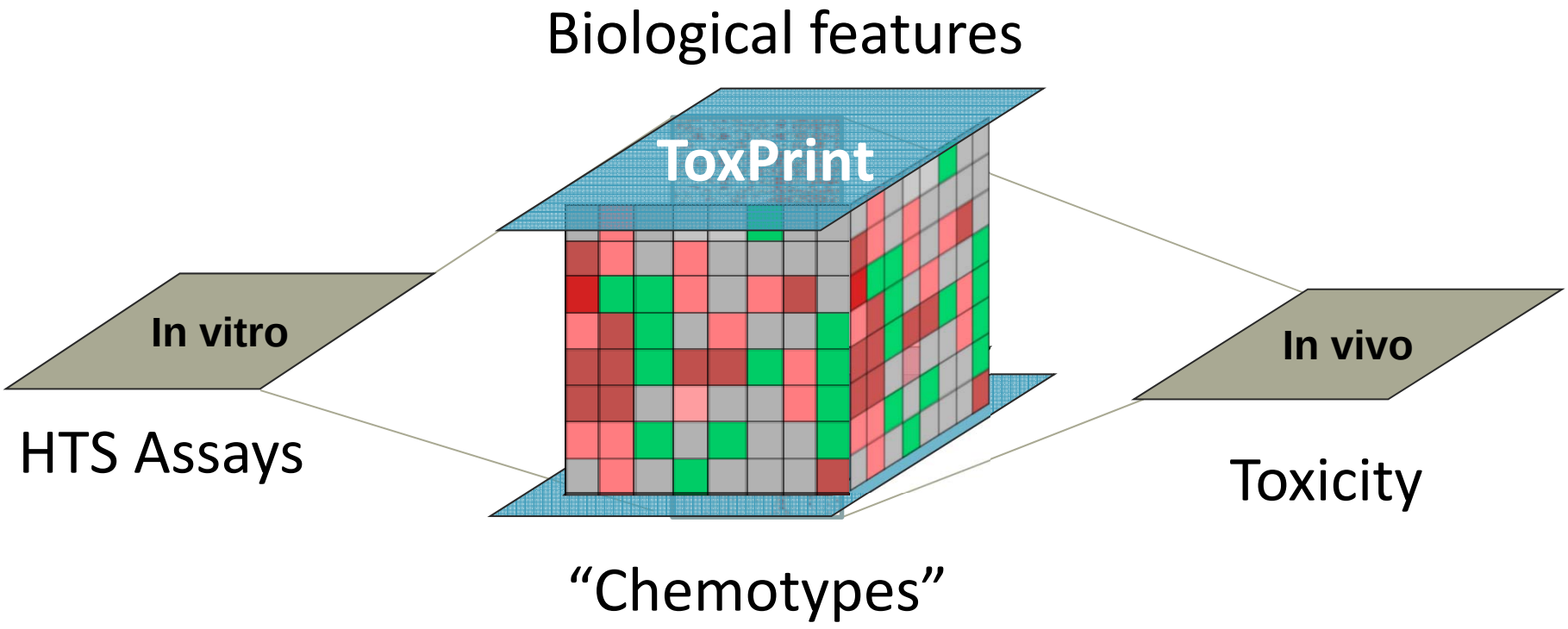
QSAR using biologically informed chemical features



HTS results are used to inform feature selection, linking chemical features, or "Chemotypes" to toxicity

Slide courtesy of C. Yang

QSAR using biologically informed chemical features



HTS results are used to inform feature selection, linking chemical features, or "Chemotypes" to toxicity

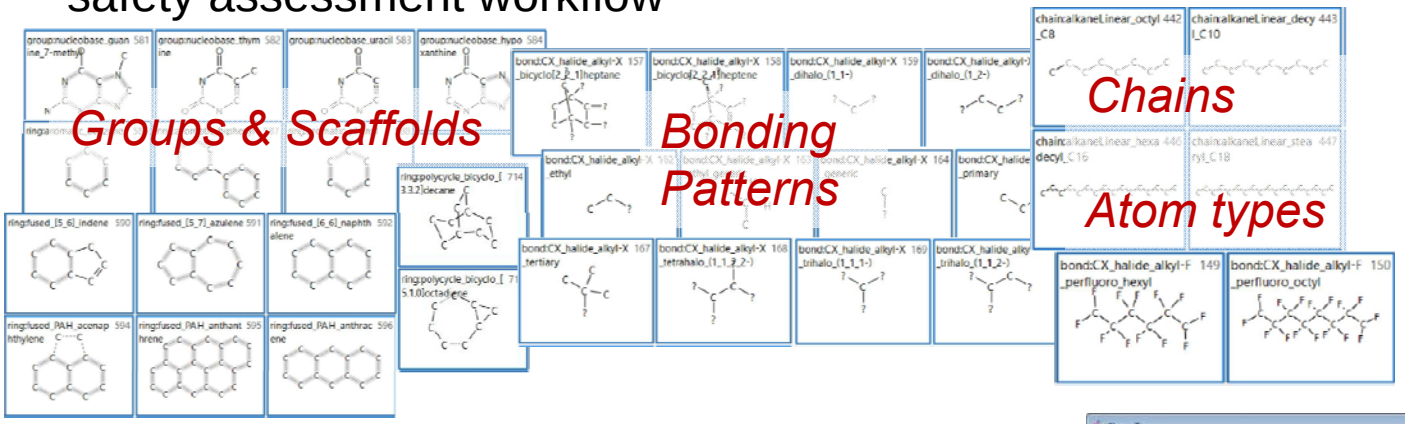
Slide courtesy of C. Yang

ToxPrints: A Public Set of Chemotypes

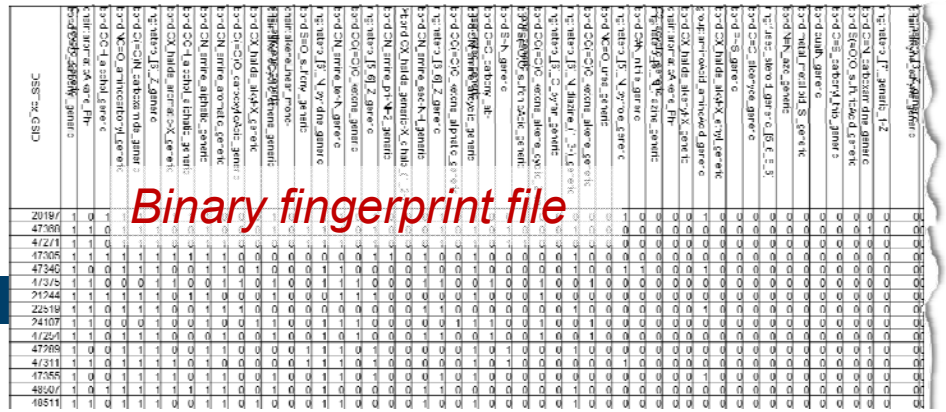
Yang, C. et al, J. *J. Chem. Inf. Model.* 55:510-28, **2015**.

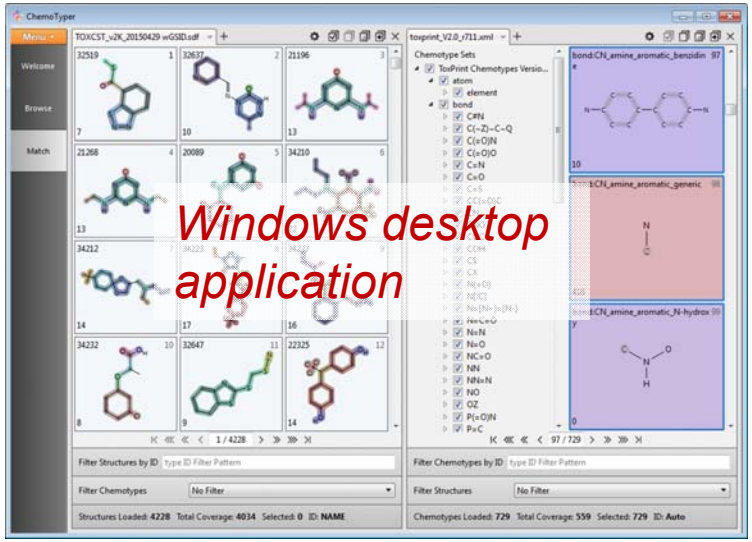
ToxPrints: http://www.toxprint.org

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

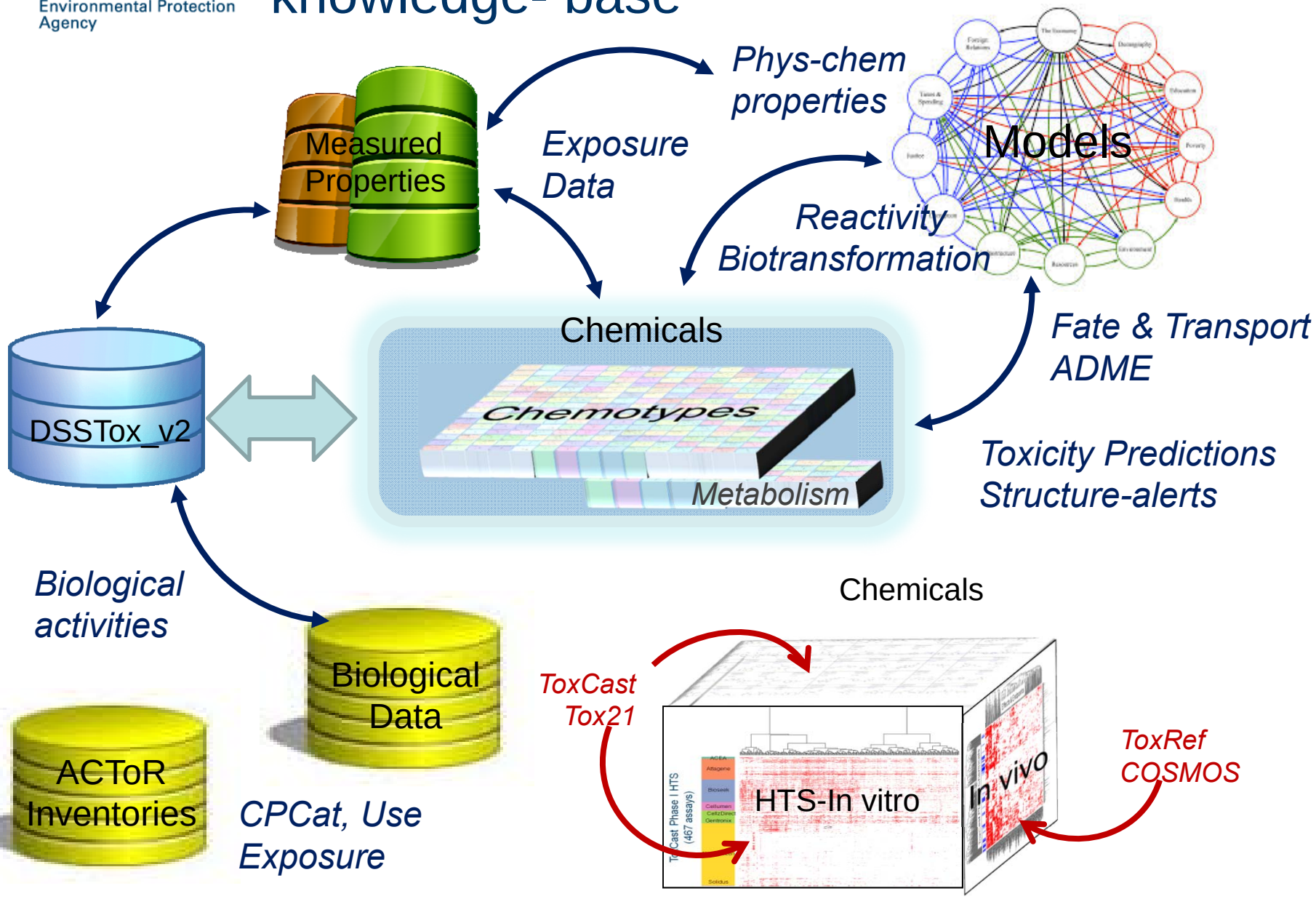


Chemotyper: http://www.chemotyper.org

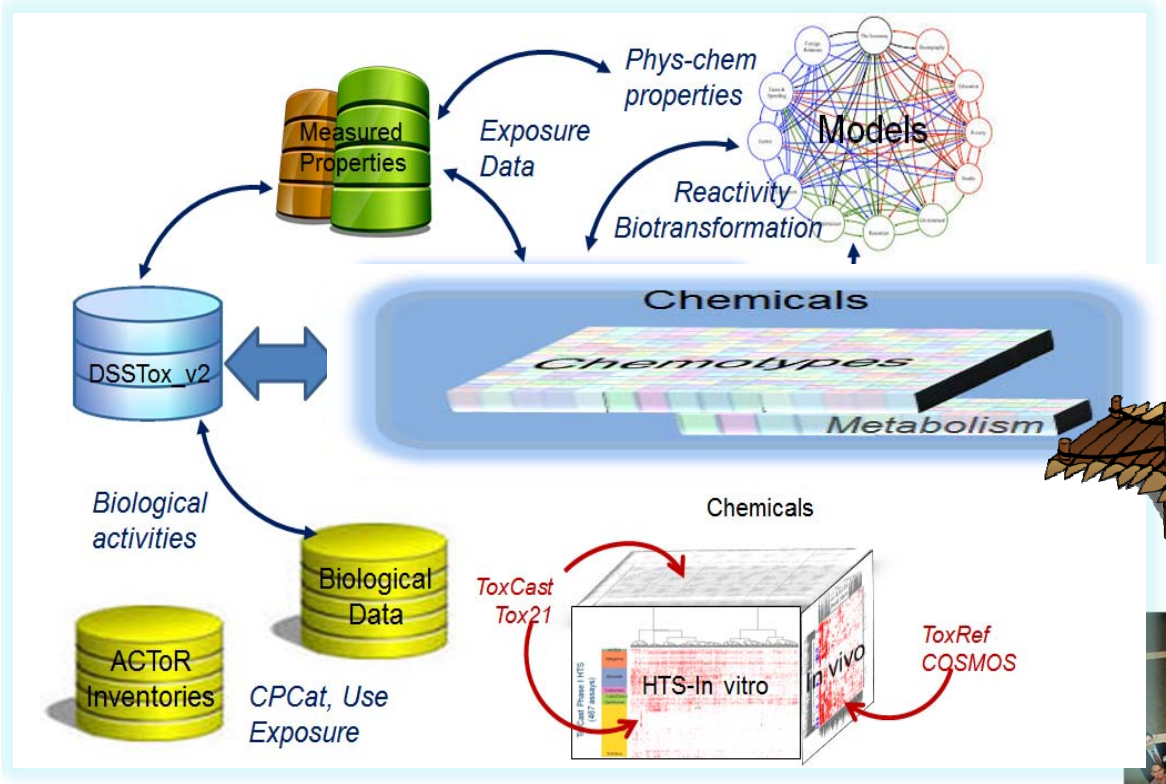


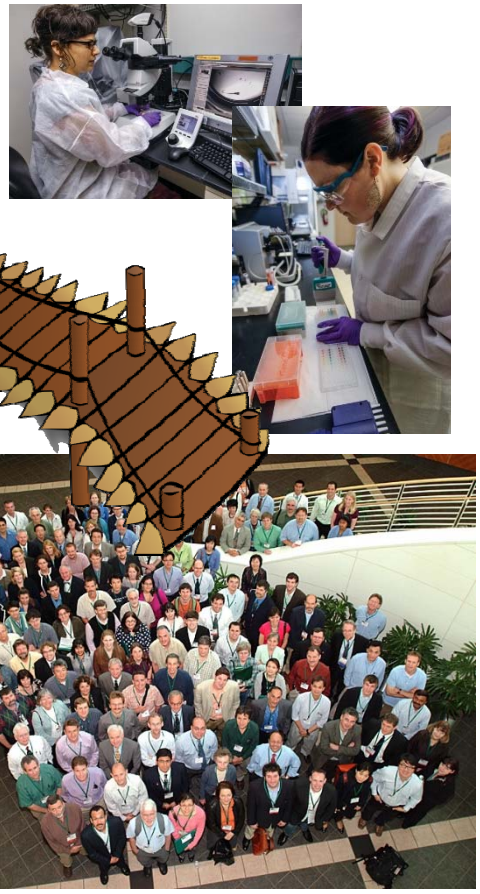


Building a public chemotype "knowledge- base"



Chemotypes: Build bridges to domain experts & knowledge resources

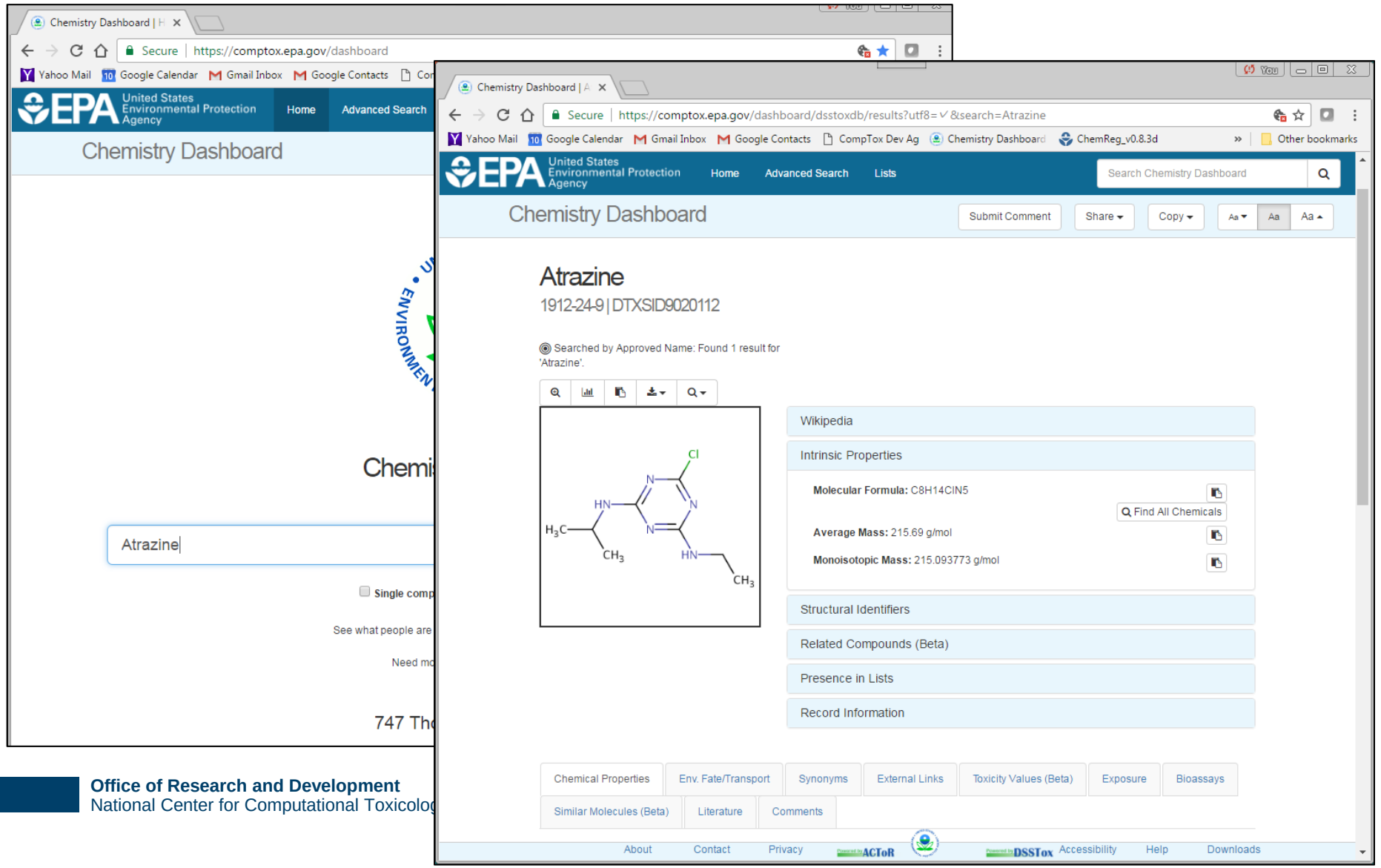




- Toxicologists
- SAR modelers
- Chemists
- Bioinformaticists
- EPA Programs
- Risk assessors

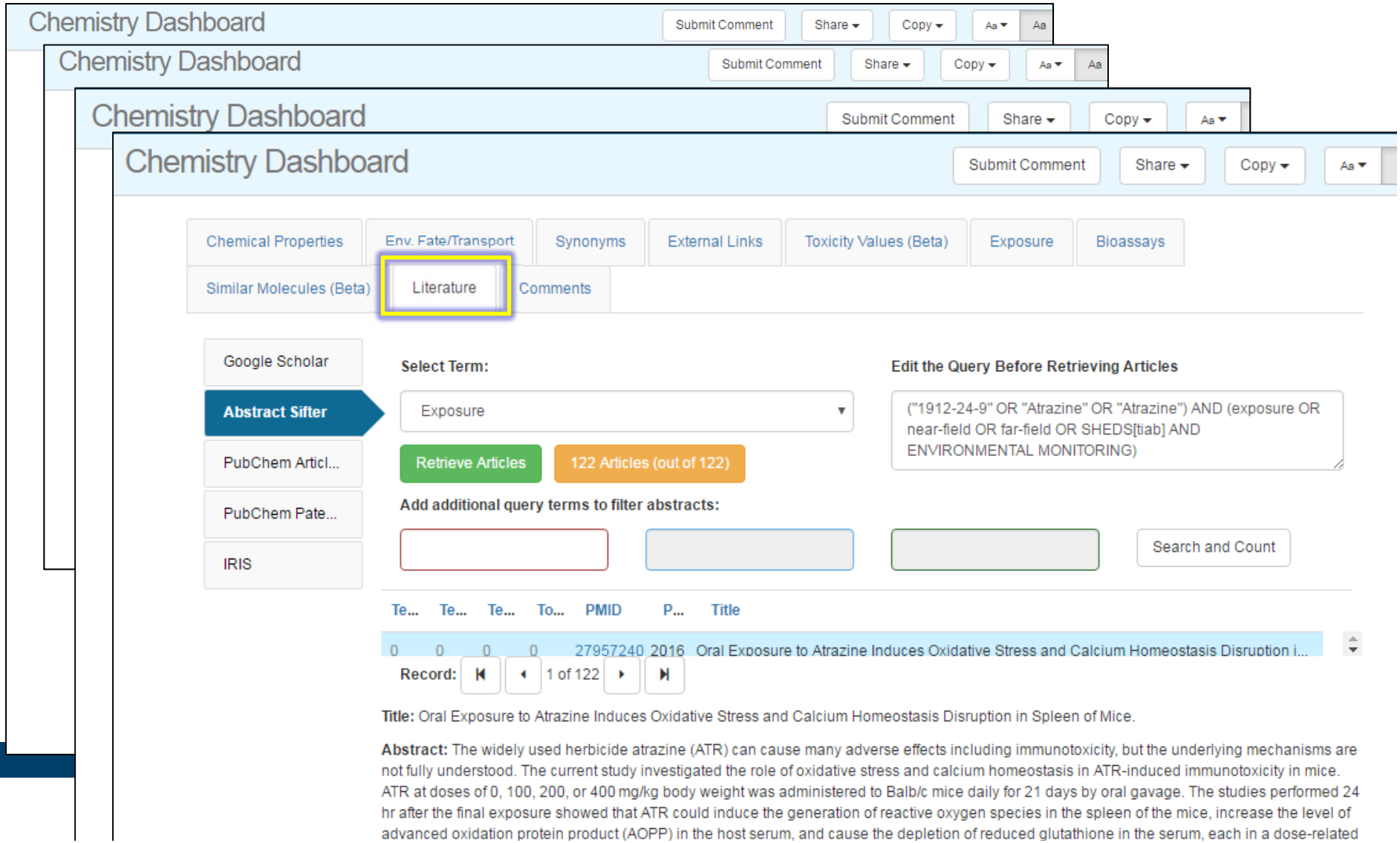
EPA's Chemistry Dashboard

https://comptox.epa.gov/dashboard



EPA's Chemistry Dashboard

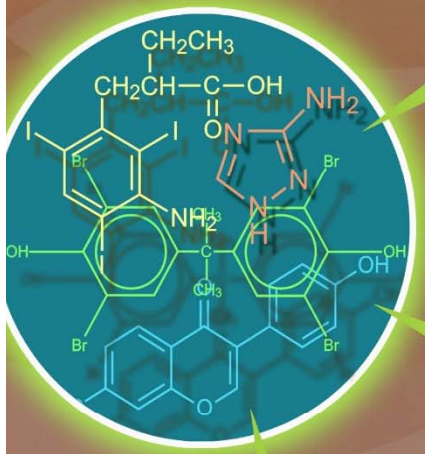
https://comptox.epa.gov/dashboard

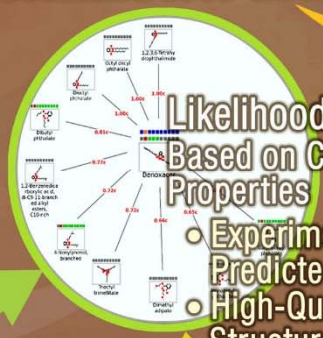


CompTox Online Dashboard

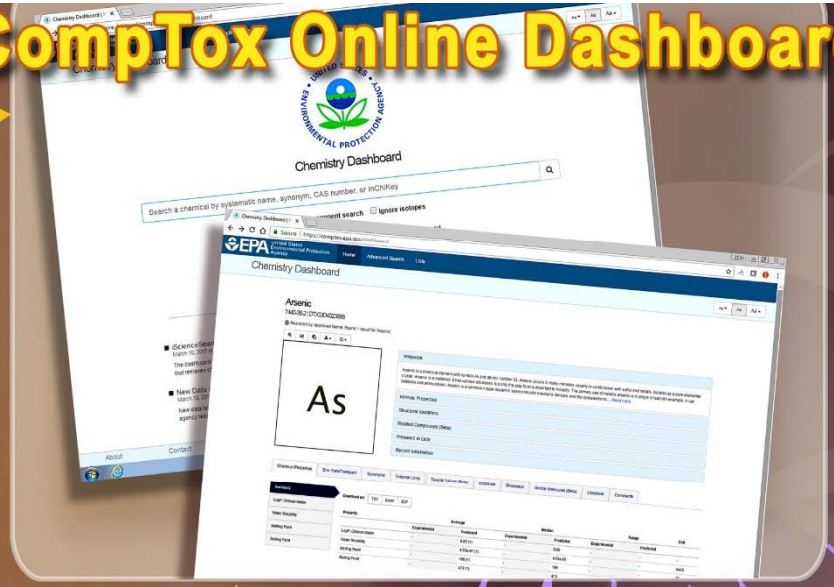
CHEMISTRY DATA

Thousands of CHEMICALS
Environment and Products





- Likelihood of Toxicity Based on Chemical Properties
  - Experimental and Predicted Values
  - High-Quality Chemical Structures



TOXICITY DATA

- High-Throughput Screening on Thousands of Chemicals and Interactions with Biological Processes
- Animal Toxicity Studies when Available

EXPOSURE DATA

- Overlay Toxicity and Exposure Data to Identify Chemicals Higher Likelihood of Potential Health Effects
  - Consumer Products
  - Predicted Total Human Exposure
  - Estimate Dose Values for Various Populations

SUPPORT TSCA ACTIVITIES

CORPORATE

CompTox Online Dashboard

Already IN USE

ENDOCRINE DISRUPTION

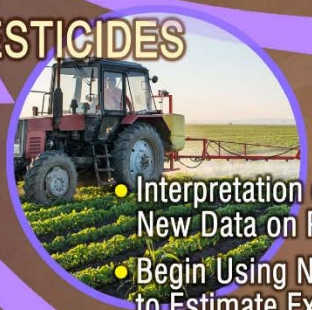
- OLD: 50-100 Chemicals per Year at Approximately $1M Each
- NEW: 1800 Chemicals in 3 Years at Approximately $30K Total

NEAR-TERM USE

CONTAMINATED SITE EVALUATION--RapidTox

- Data-Poor Chemicals Found on Site
- Integrate Related Chemical Properties Toxicity and Exposure

PESTICIDES



- Interpretation of New Data on Pesticides
- Begin Using New Models to Estimate Exposure and Dose

SUPPORT TSCA ACTIVITIES

CORPORATE

ThermoFisher SCIENTIFIC, Unilever, DOW

INTERNATIONAL

IMAP, ACCELERATED ASSESSMENT OF INDUSTRIAL CHEMICALS IN AUSTRALIA, Health Canada

STATES

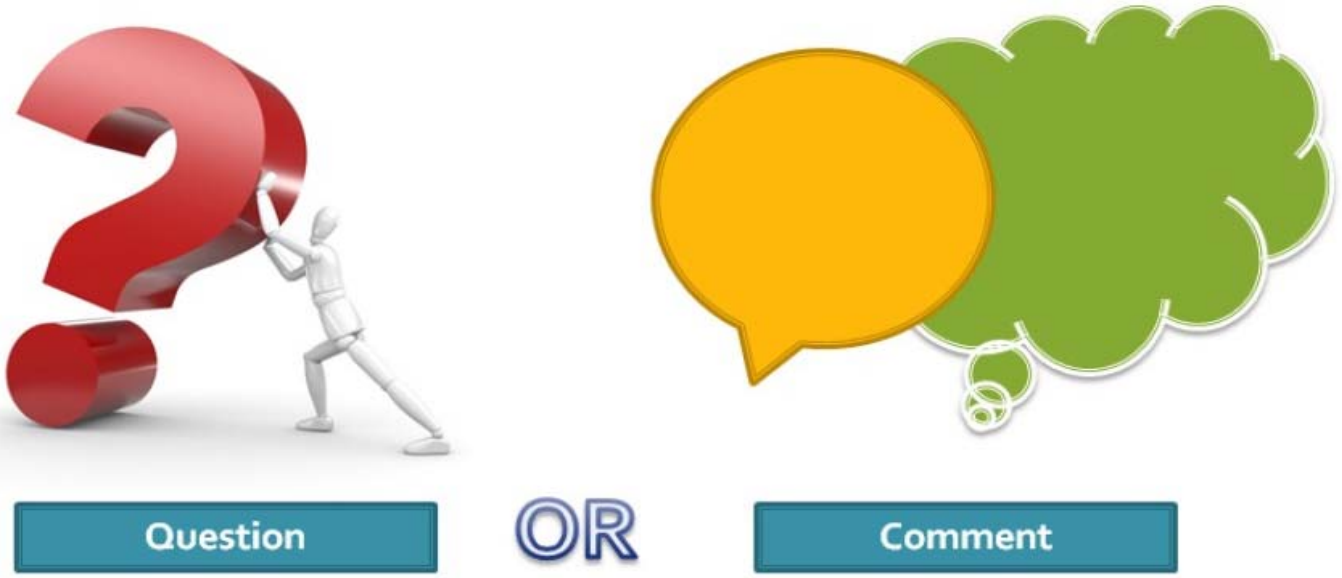
- CALIFORNIA Prioritization for Biomonitoring Pesticides
- MINNESOTA Water Contaminants

Acknowledgements:

- EPA NCCT Chemistry Team
  - Chris Grulke (DSSTox, Chemotypes)
  - Indira Thillainadarajah (DSSTox)
  - Tony Williams (DSSTox, Chemistry Dashboard)
  - Grace Patlewicz (Read-across)
- Past & present collaborators
  - John Shim
  - ClarLynda Williams-Devane
  - Chihae Yang, Jim Rathman
  - and many more…
- Entire EPA ToxCast team >

United States Environmental Protection Agency

Thank you for your attention



ToxCast Chemical Landscape

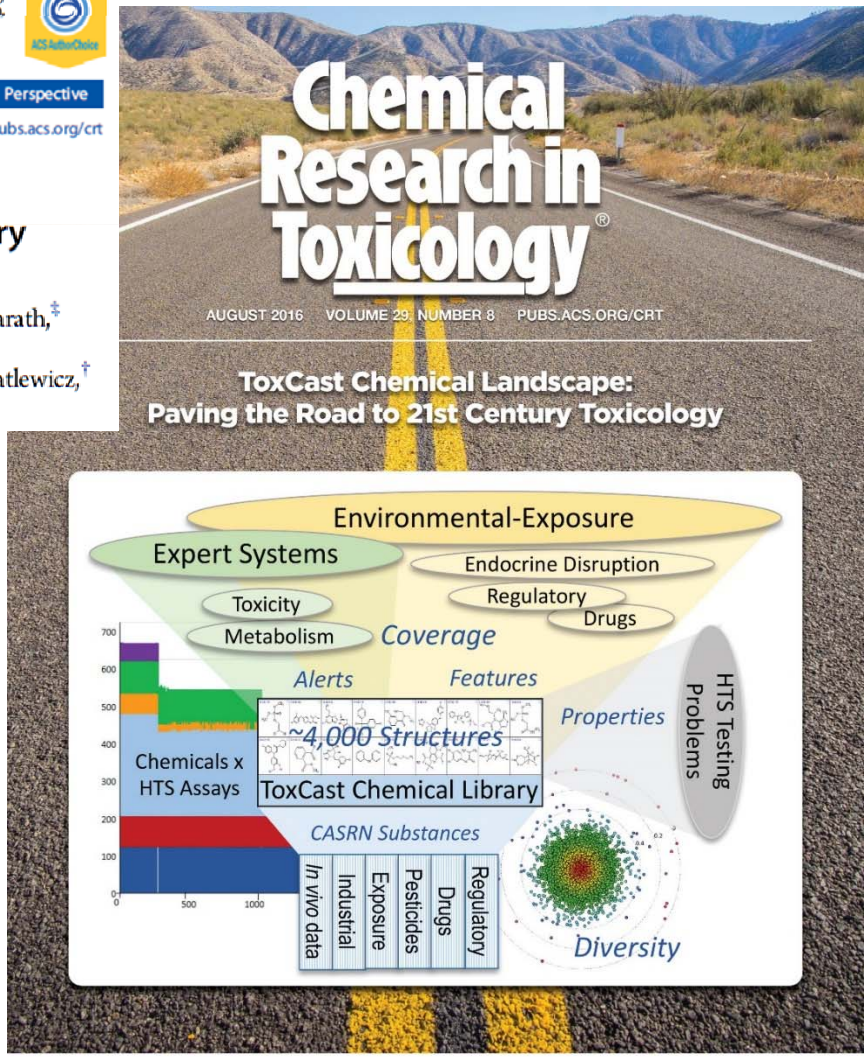
Chemical Research in Toxicology

Perspective

pubs.acs.org/crt

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

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



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

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

DOI: 10.1021/acs.chemrestox.6b00135

ChemResToxicol., 2016, 29, 1225−1251

EPA Training Opportunities for recent BA graduates

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