

Progress in Using Big Data in Chemical Toxicity Research at the National Center for Computational Toxicology

Antony Williams

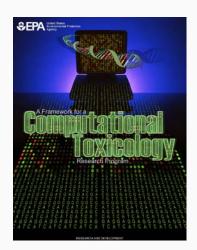
U.S. Environmental Protection Agency, RTP, NC

This work was reviewed by the U.S. EPA and approved for presentation but does not necessarily reflect official Agency policy.

June 1st, 2017 Proctor and Gamble, Cincinnati, OH

National Center for Computational Toxicology





- National Center for Computational Toxicology (2005) established to integrate:
 - High-throughput and high-content technologies
 - Modern molecular biology
 - Data mining and statistical modeling
 - Computational biology and chemistry

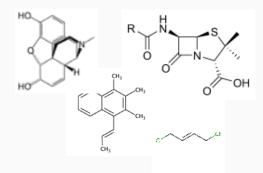


- Currently staffed by ~60 employees as part of EPA's Office of Research and Development
- Home of ToxCast & ExpoCast research efforts
- Key partner in U.S. Tox21 federal consortium

Why we must innovate...

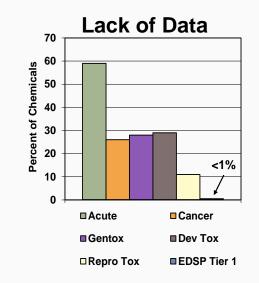


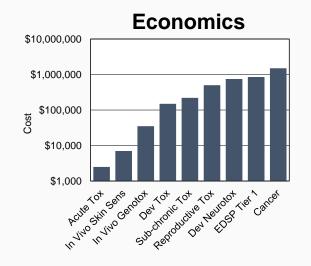
Number of Chemicals /Combinations



Ethical Concerns







Toxic Substances Control List





- Inventory was initially published in 1979
- Second version, containing about 62k chemical substances, was published in 1982
- Continues to grow and now lists ~85k chemicals, about 15k are confidential business information

The need for data and derivative models and algorithms

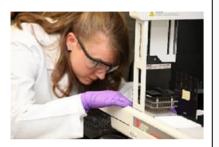


- NCCT outputs: include a lot of data, models, algorithms and software applications
- We produce Open Data we want people to interrogate it, learn from it, develop understanding

Toxicity Forecasting

Advancing the Next Generation of Chemical Evaluation

EPA needs rapid and efficient methods to prioritize, screen and evaluate thousands of chemicals. EPA's Toxicity Forecaster (ToxCast) generates data and predictive models on thousands of chemicals of interest to the EPA. ToxCast uses high-throughput screening methods and computational toxicology approaches to rank and prioritize chemicals. In fact, EPA's Endocrine Disruption



Screening Program (EDSP) is working to use ToxCast to rank and prioritize chemicals.

- ToxCast has data on over 1,800 chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially "green" chemicals that could be safer alternatives to existing chemicals.
- ToxCast screens chemicals in over 700 high-throughput assays that cover a range of high-

Downloadable Computational Toxicology Data

EPA's computational toxicology research efforts evaluate the potential health effects of thousands of chemicals. The process of evaluating potential health effects involves generating data that investigates the potential harm, or hazard of a chemical, the degree of exposure to chemicals as well as the unique chemical characteristics.

As part of EPA's commitment to share data, all of the computational toxicology data is publicly available for anyone to access and use.

High-throughput Screening Data

EPA researchers use rapid chemical screening (called high-throughput screening assays) to limit the number of laboratory animal tests while quickly and efficiently testing thousands of chemicals for potential health effects.

• ToxCast Data: High-throughput screening data on thousands of chemicals.

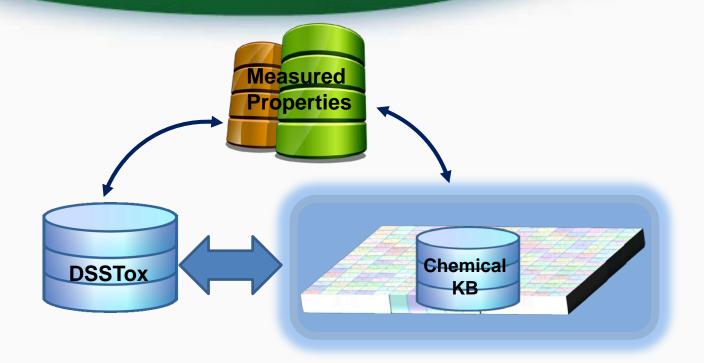
Rapid Exposure and Dose Data

EPA researchers develop and use rapid exposure estimates to predict potential exposure for thousands of chemicals.

 <u>High-throughput toxicokinetics data</u>: It is important to link the external dose of a chemical to an internal blood or tissue conceptration, this process is called toxicokinetics. EDA researchers measure the critical factors that determine the distribution.

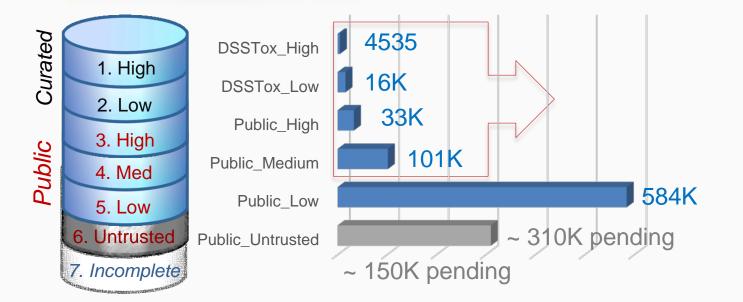
Underpinning with chemicals





Approximately 15 Years of Data... Growing with daily curation





QC Levels

DSSTox_High:	Hand curated and validated
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem
Public_Untrusted:	Postulated, but found to have conflicts in public sources

Is 750,000 substances BIG DATA?

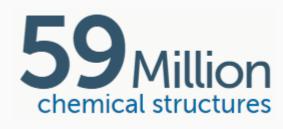


• # of chemical substances in databases

Pub©hem

Compounds: 91,490,492 Substances: 230,805,802 BioAssays: 1,252,791 Tested Compounds: 2,395,818 Tested Substances: 3,818,645 RNAi BioAssays: 112 BioActivities: 233,481,319 Protein Targets: 10,342 Gene Targets: 22,079		
	Substances: BioAssays: Tested Compounds: Tested Substances: RNAi BioAssays:	230,805,802 1,252,791 2,395,818 3,818,645 112
	BioActivities: Protein Targets:	233,481,319 10,342

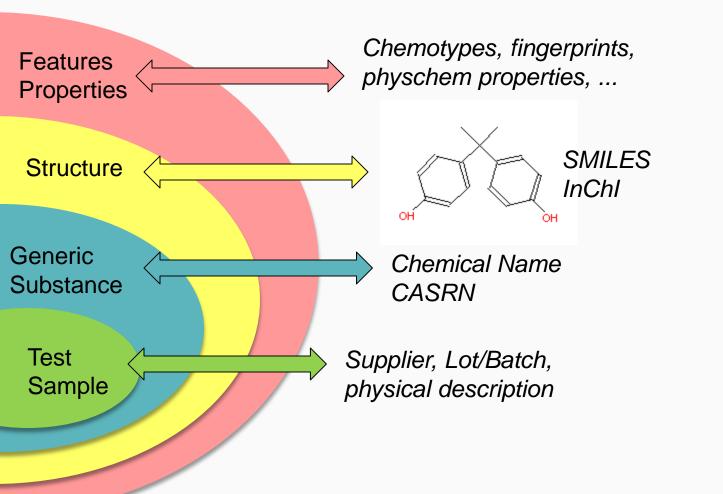
ChemSpider Search and share chemistry



E	MBL-EBI	
	🗘 ChEMBL	
	ChEMBL Statistics	
	 DB: ChEMBL_23 	
	 Targets: 11,538 	
	 Compound records: 2,101,843 	
	 Distinct compounds: 1,735,442 	

- Activities: 14,675,320
- Publications: 67,722
- Release Notes
- Data quality is a challenge!!! Careful curation

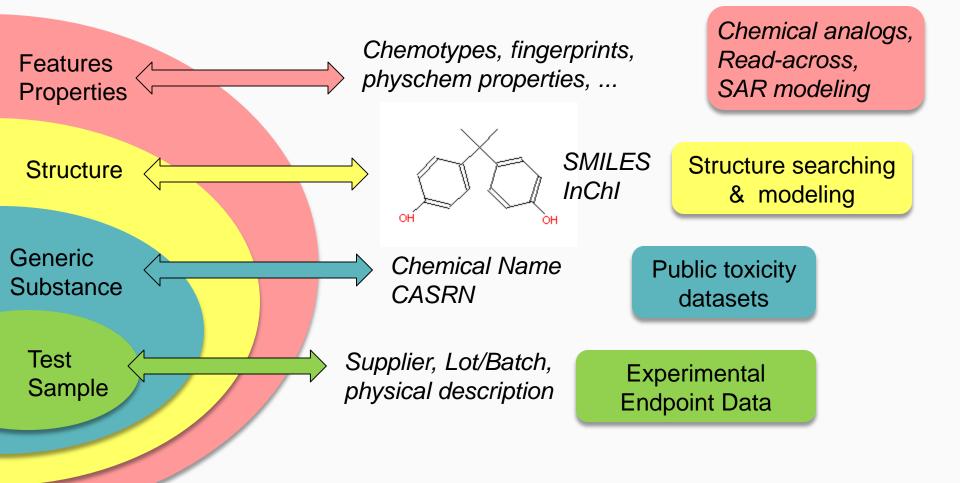
Chemical representation levels supporting data integration



imental Protection

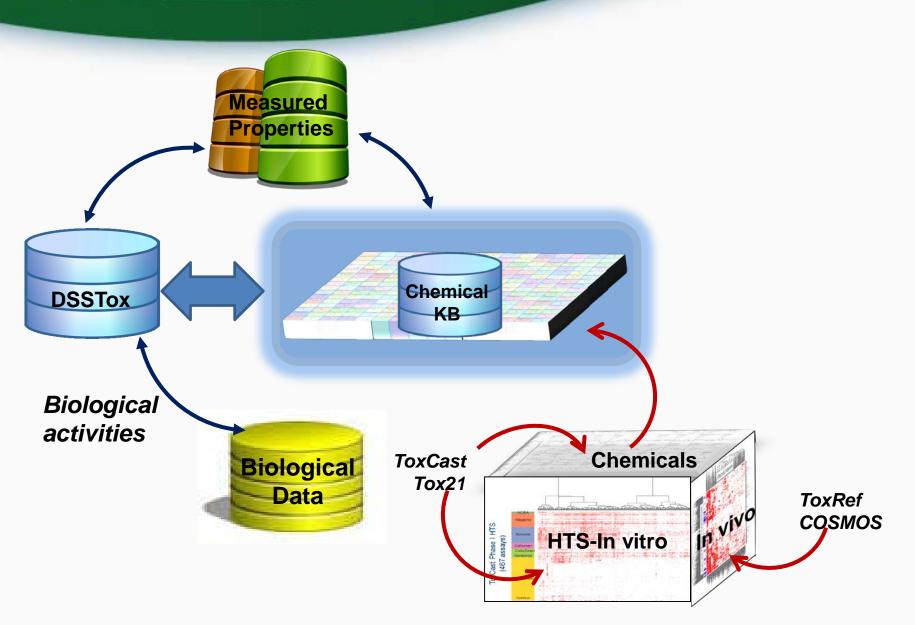
Chemical representation levels supporting data integration





9

Integrating in vitro and in vivo data



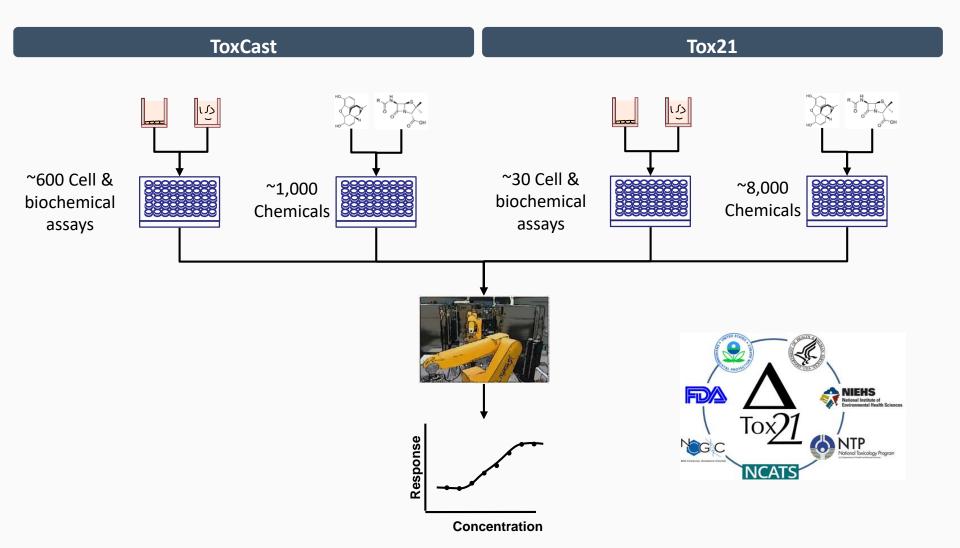
€EP

Agency

wironmental Protection

Adding High-Throughput Hazard Screening Data





TEN YEARS of Assay Measurements: ToxCast & Tox21



Set	Chemicals	Assays	Endpoints	Completion
ToxCast Phase I	293	~600	~700	2011
ToxCast Phase II	767	~600	~700	03/2013
ToxCast E1K	800	~50	~120	03/2013
ToxCast Phase III	~900	~300	~300	In progress
Tox21	~9000	~80	~150	In progress

Pesticides , antimicrobials, food additives, green alternatives, HPV, MPV, endocrine reference cmpds, tox reference cmpds, NTP in vivo, FDA GRAS, FDA PAFA, EDSP, water contaminants, exposure data, industrial, failed drugs, marketed drugs, fragrances, flame retardants, etc.

Chemicals

~800

Assays

FD/A

GC

NCATS

 $\mathbf{0}$

NIEHS

NTP

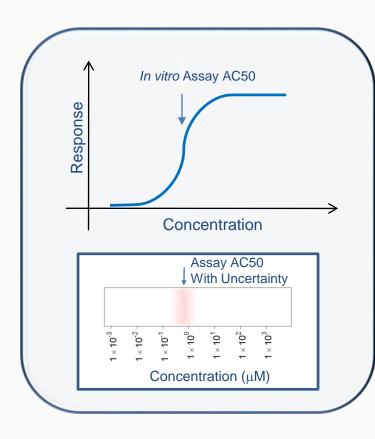


~9000

High-Throughput Bioactivity to Identify **Potential Hazards**

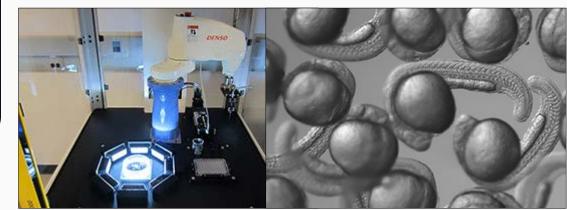


Assays in dose-response format (AC50 – and efficacy if data described by a Hill function)

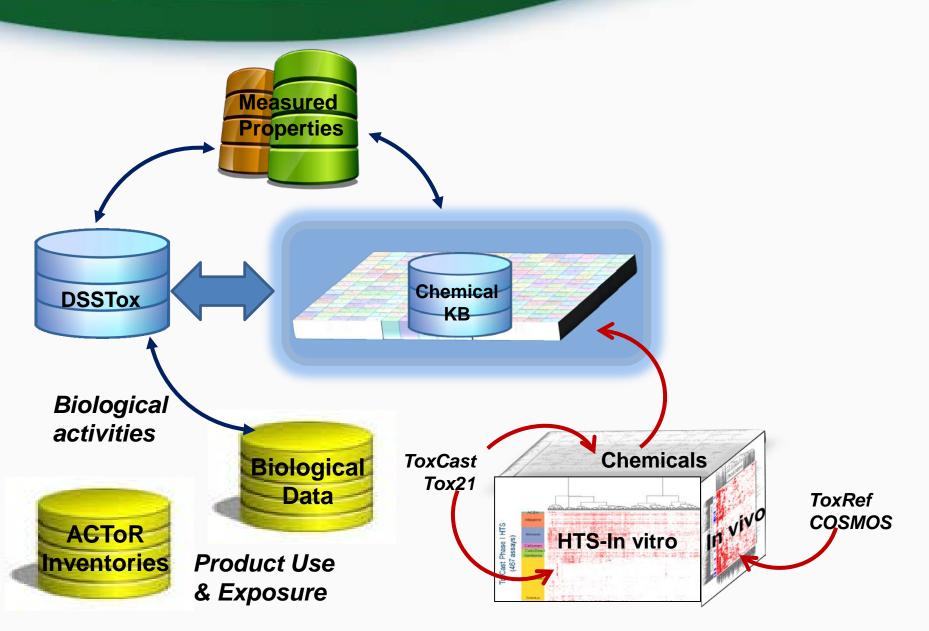


Choose a view:	 Assays Chemicals 	Chemical Summary Anag Summary Bloodbilly Help Bart Marial - Cemical Sa			
Chemicals - 10			Assays - 1011 0, 🗸 🤀	Chemical Autority Summary	
CASRN.	bispherol A		Assay Endpoint Name	See Dart Preview Dart	
CASRN	Chemical Name	Che	Actives - MC Only As Tested		
19-05-8	2.2.0.5-Tetrachtorobisphenol A	phen	Assay Component Endpoint Name Gene Symbol a AGEA_T47D_SDV: Negative A	Active endpoints for ID-05-7 Bacilground measurement P.7 B cell adhesion molecules	
19-94-7	1.7.5.5-Tetrabromobilsphendi A	phen	# ACEA_TATO_BOY_POSTNE ESR1	e O nel syste	
79-97-0	3.3-Olmethylbisphenol A		# APR_Hep02_Cel0ydeAmest_th_dn	7.8 · · · · · · · · · · · · · · · · · · ·	
10-05-7	Baphensi A	phen	# APR_HepG2_CellOysteArrest_Th_up		
1075-54-3	Bispherol A dgycogi ether Bispherol A dgycogi metheoviate	1940	H APR_Hep32_Celluss_Ih_ge H APR_Hep32_Celluss_Ih_ge H APR_Hep32_Celluss_Ih_ge H APR_Hep32_NormaniadeCSK_Ih_ge	e nyo O cisi B B B B B B B B B B B B B B B B B B	
478-61-1	Bischend A gyody metheoryate Bischend AF	phen	W APR_Hep02_MonstableC5K_In_up # APR_Hep02_MonMass_In_dn	esterase	
102-45-2	Setabornobisphenol A bis(2-hydroxyethyl) ether		# APR,Hep02,MtoMeet_th_se # APR,Hep02,MtoMeetCh_pt_an	221 23 23 25 25	
5927-89-3	Tetrabromobiaphenol A diallyl ether	_	* APR Hes02 MinMemiPol To up	10	
21800-44-2	Tetrabiomobisphenol A-bis(2,3-dohomopropy) ether)		APR Hard's Management (n. de	00 el con channel	
itters - 0			00	0001 0,01 0,11 AC50 (uM) 10 100 100 €knase ⊕lysse	
ist	Field	Va	ue	Scaled importance is calculated by dividing the importance values by the activity scale in scale activity scale	

All data is made public: http://actor.epa.gov/dashboard/



Adding Product Use and Exposure



€EP

Anenity

vironmental Protection

High Throughput Measurement to Identify **Exposure**



	vironmental Protection Agency	Esp	añol	中文: 繁體版		中文: 简体版 丨	Tiếng Việt	한국어	
Learn the Issues	Science & Technology	Laws & Regulations	About EPA				Search EPA.g	ov	٩
Related Topics: Safer Chemicals Research Contact Us									

Rapid Chemical Exposure and Dose Research

EPA is responsible for ensuring the safety of thousands of chemicals. Quantitative exposure data are available for only a small fraction of registered chemicals. This type of exposure data is needed to thoroughly evaluate chemicals for potential risks to humans, wildlife and ecosystems. EPA is developing innovative methods to develop exposure estimates for thousands of chemicals to better protect human health and the environment. These innovative methods are called rapid exposure and dose assessments.

Rapid Exposure Predictions

Rapid, also called high-throughput, exposure predictions or ExpoCast provide rapid exposure estimates for thousands of chemicals. ExpoCast quickly and efficiently looks at multiple routes of exposure to provide exposure estimates. ExpoCast uses and enhances two well-known exposure models to estimate chemical exposure.

- > Farfield Exposure Models
- > Nearfield Exposure Models

Evaluating High-throughput Exposure Predictions

EDA is currently avaluating the effectiveness of high throughout evenesure models



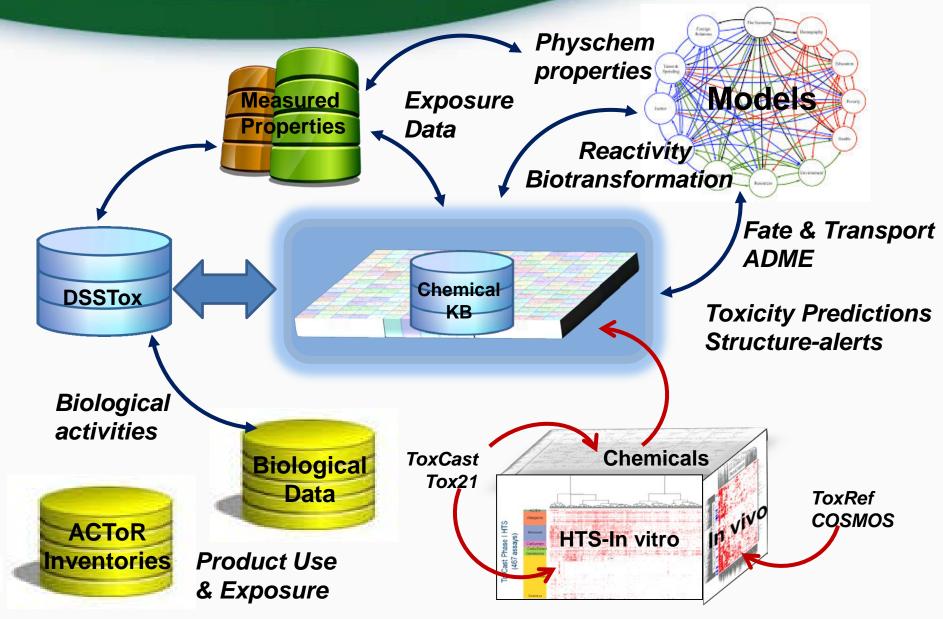
Pictured Above: Farfield Exposure Examples



antal Pratactics

Building Models from the data

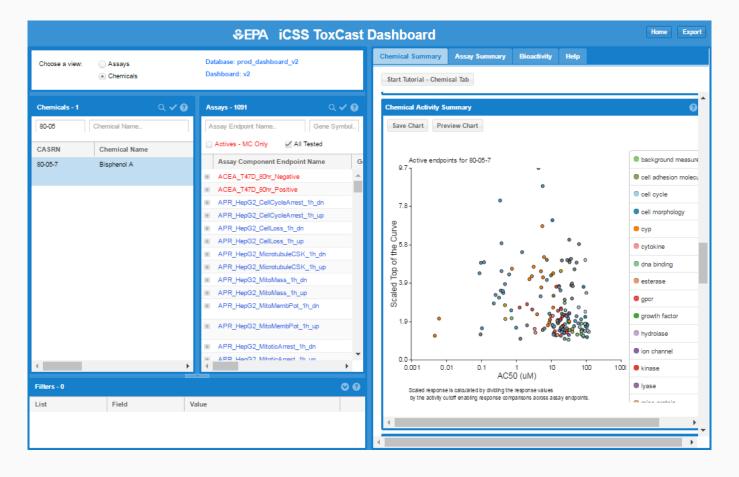




Toxcast Dashboard https://actor.epa.gov/dashboard/

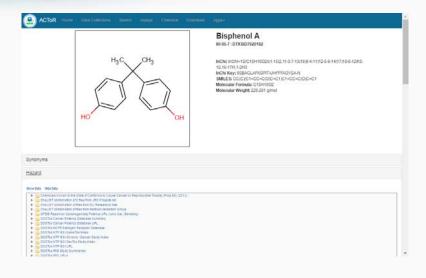


 Access and Interrogate chemical screening data from ToxCast and the Tox21 collaboration

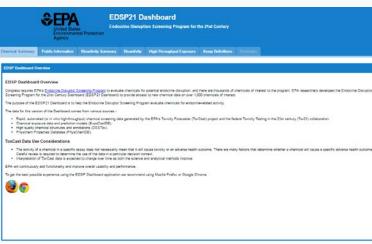


Multiple Dashboards Delivered





EPA United States						
Cat: Chemical and		gories cology Research - Chemica	i Use		EEEContact Us	
Ditome Disearch	* Results	a Dictionary & Downlo	and Allelp			
Chemical: BISPHENOL	A					
	C/	SRN: 80-05-7				
Export Use Data	DH Export Produ					
lse Informations						
CPCat Descrip	ption ¢	Source Description o	ACTOR Data Set/List ©	Scurce q	Class of Chemical Category	
consumer_use_ACToR1	JseDB	Consumer Use		ACTOR UseDB	Use Categories	
personal_care_ACToRU	lseOB	Personal Care Product		ACToR UseDB	Use Categories	
industrial_manufacturing	ACTORUSEDS	Chemical Industrial		ACTOR UseOB	Use Categories	
shid_use detected		Consumer Products	The Danish EPK Exposure of 2-year-obts to chemical substances in Consumer Products. This project included a survey of the products as well as chemical analyses and risk assessments of a number of	ACToR Data Sets and Usis	Use Categories	

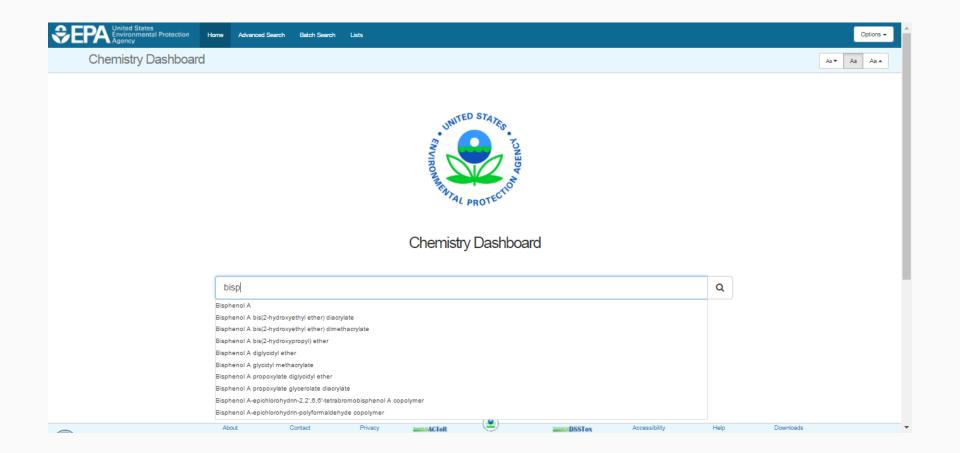




- Data curation, standardization and versioning is essential
- Prototype application development suffices for research projects but production development requires managed processes
- ODOSOS (Open Data, Open Source and Open Standards) endows many benefits
- With this in mind...

Our Latest Dashboard https://comptox.epa.gov





Bisphenol A



United States Environmental Protection Agency	Home Advanced S	earch Batch Sea	rch Lists					Search Chemistr	y Dashboard	
emistry Dashboa	rd						Submi	Comment	Share 👻	Сору 🕶
Bisphenol A 80457 DTXSID7020182										
	Name: Found 1 result for 'b	isphenol A'.								
ର୍ ଲା 🖪 🛓	T ((T	7	Wikipedia							
H ₃ C	CH ₃		diphenylmeth soluble in wa	ane derivatives and bisp	henols, with two	hydroxyphenyl g	l formula (CH3)2C(C6H4O) groups. It is a colorless solid ployed to make certain plast	that is soluble in o	organic solvents	
	>		Intrinsic Pro	operties						
			Structural lo	dentifiers						
но	OF	l l	Related Co	mpounds (Beta)						
			Presence in	n Lists						
L			Record Info	ormation						
Chemical Properties	Env. Fate/Transport	Synonyms Ex	ernal Links	Toxicity Values (Beta)	Exposure	Bioassays	Similar Molecules (Beta	Literature	Comment	ts

_

Contact

About

ACTOR

Privacy

۲

Accessibility

DSSTox

Help

Downloads

Chemical Properties

About

Env. Fate/Transport

Chemical Properties

Contact

Synonyms

Privacy

External Links



United States Environmental Protection Agency	Home Advanced S	earch Batch Searc	:h Lists					Search Chemistr	y Dashboard	Q
emistry Dashboa	rd						Subr	nit Comment	Share 🗸 🛛 🔾	Copy 🔻
Bisphenol A 80-057 (DTXSID7020182										
Searched by Approved Q Im Im ±	Name: Found 1 result for 'bi	isphenol A'.								
			Wikipedi	a						
H ₃ C	CH ₃		diphenylm soluble in	A (BPA) is an organic synth lethane derivatives and bisp water. It has been in comm tough Read more	henols, with two	hydroxyphenyl g	roups. It is a colorless so	d that is soluble in o	rganic solvents, bu	
	>		Intrinsic I	Properties						
			Structural Identifiers							
но	ОН	I	Related Compounds (Beta)							
			Presence	e in Lists						
		-	Record I	nformation						
Chemical Properties	Env. Fate/Transport	Synonyms Exte	rnal Links	Toxicity Values (Beta)	Exposure	Bioassays	Similar Molecules (Be	a) Literature	Comments	

۲

Toxicity Values (Beta)

DSSTox

Exposure

Accessibility

Bioassays

Help

Similar Molecules (Beta)

Downloads

Literature

Comments

****ACToR

Chemical Properties



Consuming and producing open data •

ogP: Octanol-Water								
-	Property	Av	erage	Me	Range	Unit		
/ater Solubility		Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
Density	LogP: Octanol-Water	3.32 (1)	3.24 (4)	3.32	3.24	3.32	2.40 to 3.73	-
	Water Solubility	5.26e-04 (1)	1.58e-03 (4)	5.26e-04	1.58e-03	5.26e-04	5.70e-04 to 3.68e-03	mol/L
elting Point	Density	-	1.14 (1)	-	1.14	-	-	g/cm^3
oiling Point	Melting Point	155 (7)	144 (3)	156	144	153 to 158	132 to 157	°C
oung oun	Boiling Point		349 (3)	200	349	200	334 to 364	°C
urface Tension	Surface Tension	-	46.0 (1)	-	46.0	-	-	dyn/cm
apor Pressure	Vapor Pressure	-	2.52e-07 (3)	-	2.52e-07	-	7.01e-08 to 5.34e-07	mmHg
	LogKoa: Octanol-Air	-	8.38 (1)	-	8.38	-	-	-
ogKoa: Octanol-Air	Henry's Law	-	6.96e-07 (1)	-	6.96e-07	-	-	atm-m3/mole
	Index of Refraction	-	1.60 (1)	-	1.60	-	-	-
enry's Law	Molar Refractivity	-	68.2 (1)	-	68.2	-	-	cm^3
dex of Refraction	pKa Acidic Apparent	-	10.3 (1)	-	10.3	-	-	-
laine Dafaashi iku	Molar ∀olume	-	200 (1)	-	200	-	-	cm^3
Molar Refractivity	Polarizability	-	27.0 (1)	-	27.0	-		Å^3

pKa Acidic Apparent

Chemical Properties

Comments

Data Distribution



Consuming and producing open data

Summary	Download as: TSV	Excel SDF						
LogP: Octanol-Water	Property	Select/Deselect A		Ме	dian		Range	Unit
Nater Solubility		 LogP: Octanol-Wa Water Solubility 	ter	xperimental	Predicted	Experimental	Predicted	
Density	LogP: Octanol-Water	Density		.32	3.24	3.32	2.40 to 3.73	-
-criaty	Water Solubility	Melting Point		26e-04	1.58e-03	5.26e-04	5.70e-04 to 3.68e-03	mol/L
lelting Point	Density	 Boiling Point Surface Tension 			1.14	-	-	g/cm^3
oiling Point	Melting Point	 Vapor Pressure 		56	144	153 to 158	132 to 157	°C
	Boiling Point	LogKoa: Octanol-A	Air	00	349	200	334 to 364	°C
urface Tension	Surface Tension	Henry's Law	_		46.0	-	-	dyn/cm
apor Pressure	Vapor Pressure	 Index of Refraction Molar Refractivity 	1		2.52e-07	-	7.01e-08 to 5.34e-07	mmHg
	LogKoa: Octanol-Air	PKa Acidic Appare	nt		8.38	-	-	-
ogKoa: Octanol-Air	Henry's Law	Molar Volume			6.96e-07	-	-	atm-m3/mole
learn de Laur	Index of Refraction	Polarizability	1		1.60	-	-	-
lenry's Law	Molar Refractivity	Downloa	ad		68.2	-		cm^3
ndex of Refraction	pKa Acidic Apparent	- 1	0.3 (1)	-	10.3	-	-	-
	Molar Volume	- 2	200 (1)	-	200	-	-	cm^3
Molar Refractivity	Polarizability	- 2	- 27.0 (1)		27.0	-	-	Å^3

pKa Acidic Apparent

Chemical Properties

Comments

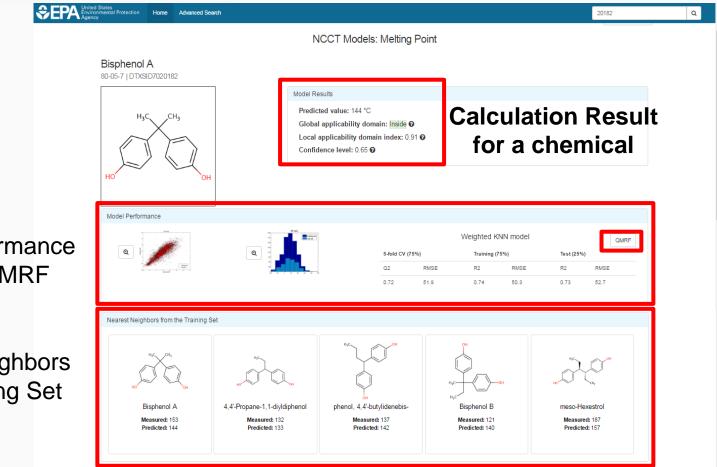
Data Distribution



	A	В	С	D	E	F	G	H	
1	Property	Avera	ige	Med	ian		Range		
2		Experimental	Predicted	Experimental	Predicted	Experimental	Predicted		
3	LogP: Octanol-Water	3.32 (1)	3.24 (4)	3.32	3.24	3.32	2.40 to 3.73	-	
4	Water Solubility	5.26e-04 (1)	1.58e-03 (4)	0.000526	0.00158	0.000526	5.70e-04 to 3.68e-03	mol/L	
5	Density	-	1.14 (1)	-	1.14	-	-	g/cm^3	
6	Melting Point	155 (7)	144 (3)	156	144	153 to 158	132 to 157	°C	
7	Boiling Point	200 (1)	349 (3)	200	349	200	334 to 364	°C	
8	Surface Tension	-	46.0 (1)	-	46	-	-	dyn/cm	
9	Vapor Pressure	-	2.52e-07 (3)	-	0.00000252	-	7.01e-08 to 5.34e-07	mmHg	
10	LogKoa: Octanol-Air	-	8.38 (1)	-	8.38	-	-	-	
11	Henry's Law	-	6.96e-07 (1)	-	0.00000696	-	-	atm-m3/mole	
12	Index of Refraction	-	1.60 (1)	-	1.6	-	-	-	
13	Molar Refractivity	-	68.2 (1)	-	68.2	-	-	cm^3	
14	pKa Acidic Apparent	-	10.3 (1)	-	10.3	-	-	-	
15	Molar Volume	-	200 (1)	-	200	-	-	cm^3	
16	Polarizability	-	27.0 (1)	-	27	-	-	Å^3	
17									

Modeling Details





Model Performance with full QMRF

Nearest Neighbors from Training Set

Prediction Details and QMRF Report



vlodel Results			
Predicted value: 144 °C Global applicability domain: Inside Local applicability domain index: 0	details in QMRF.		
Confidence level: 0.65 @			
☑ QMRF_NCCT_MP_08212016 - Add File Edit View Window Help	be Acrobat Pro		
🔁 Create 🗸 📑		Customiz	e •
	Tools	Fill & Sign	Comn
■ <i>■</i> × <i>×</i>	QMRF identifier (JRC Inventory): To be entered by JRC QMRF Title: MP: Melting point prediction from the NCCT Models Suite. Printing Date: May 4, 2016		
	1.QSAR identifier		
	1.1.QSAR identifier (title): MP: Melting point prediction		
	from the NCCT_Models Suite.		
	1.2.Other related models:		
	No related models		
	1.3.Software coding the model:		
	NCCT_models V1.02		
	Suite of QSAR models to predict physicochemical properties and environmental fate of organi chemicals	с	

Developing "NCCT Models"

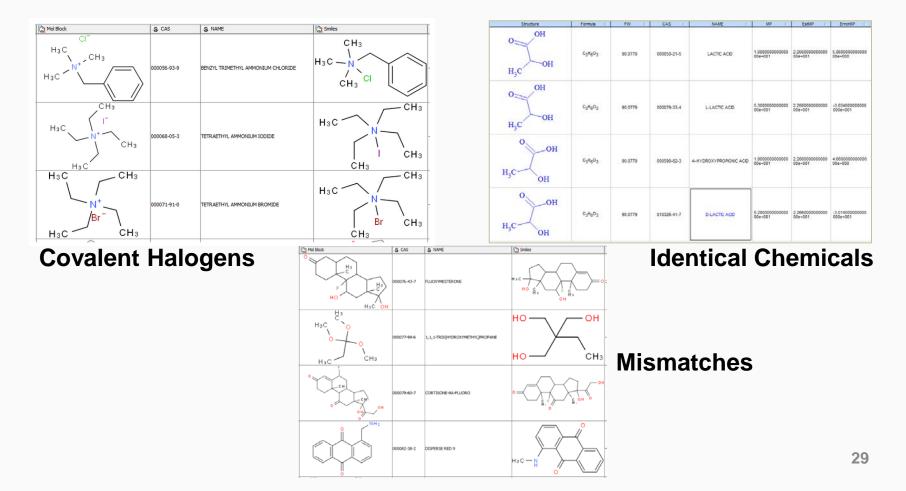


- Our approach to modeling:
 - Obtain high quality training sets
 - Apply appropriate modeling approaches
 - Validate performance of models
 - Define the applicability domain and model limitations
 - Use models to predict properties across our full datasets
 - Release as Open Data and Open Models

Consuming and Curating Public Data

CEPA

Public data should be curated prior to modeling



- CAS Checksum: 12163 valid, 3646 invalid (>23%)
- Invalid names: 555
- Invalid SMILES 133
- Valence errors: 322 Molfile, 3782 SMILES (>24%)
- Duplicates check:
 - -31 DUPLICATE MOLFILES
 - -626 DUPLICATE SMILES
 - -531 DUPLICATE NAMES
- SMILES vs. Molfiles (structure check)
 - -1279 differ in stereochemistry (~8%)
 - -362 "Covalent Halogens"
 - -191 differ as tautomers
 - -436 are different compounds (~3%)

Workflow Details and Data

lournal

SAR and QSAR in Environmental Research >

Articles

Volume 27, 2016 - Issue 11: 17th International Conference on QSAR in Environmental and Health Sciences (QSAR 2016) - Part II. Guest Editors: C.G. Barber and G.J. Myatt

Enter keywords, authors, DOI etc.

258 Views 4 CrossRef citations 16 Altmetric

An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling^{\$}

Check for updates

K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams 🔤 Pages 911-937 | Received 03 Sep 2016, Accepted 24 Oct 2016, Published online: 25 Nov 2016

http://dx.doi.org/10.1080/1062936X.2016.1253611 66 Download citation

OPERA Models: https://github.com/kmansouri/OPERA



Synonyms – curated and Open



Chemical Properties	Env. Fate/Transport	Synonyms	External Links	Toxicity Values (Beta)	Exposure	Bioassays	Similar Molecules (Beta)	Literature	Comments	
				Found 141 sy	/nonyms					
		Legend:	Valid Synonym	s Good Synonyms	Other Synonyms	Copy a	all Synonyms			
Bisphenol A										
4,4'-(Propane-2,2-diyl)d phenol, 4,4'-(1-methyle	-									
80-05-7 Active CAS-RN										
BPA 4,4'-Propane-2,2-diyldij	phenol									- 1
Phenol, 4,4'-(1-methyle	thylidene)bis-									
4-06-00-06717 Belistein Re										
(4,4'-Dihydroxydiphenyl)d 2,2-Bis(4'-hydroxyphenyl)	-									
2,2'-Bis(4-hydroxyphenyl) 2,2'-Bis(4-hydroxyphenyl)										
2,2-BIS-(4-HYDROXY-PH										
2,2-Bis(4-hydroxyphenyl)	propane				<u>_</u>					

Taking advantage of online data



Chemical Properties	Env. Fate/Transport	Synonyms External Links	Toxicity Values (Beta)	Exposure	Bioassays	Similar Molecules (Beta)	Literature	Comments	
General	Toxicology	Publicatio	ons A	nalytical		Prediction			
EPA Substance Regis	st 🖲 ACToR	Toxline		RSC Analytical	Abstra	oo Chemicalize			
eChemPortal p	provides free public acce	ss to information on propert	ties of	FOR-IDENT		Proton NMR Prediction			
🛷 H chemicals. Dire	ect links to collections of	f chemical hazard and risk	Pr 0	C MONA: MassBank No		Carbon-13 NMR Pred			
	epared for government ch nal and international leve	hemical review programmes els are obtained.	at	8 NEMI: National	I Enviro	C 2D NMR HSQC/HMB			
💢 d		•		Tox21 Analytics	al Data	ChemRTP Predictor			
hmp HMDB	Section 2015 Section 2015	al Google	e Patents						
w Wikipedia	() EDSP Dashl	board DubMe	ed						
Q MSDS Lookup	Gene-Tox	Q BioCad	ddie DataMed						
Q ToxPlanet		Q Federa	al Register						
Q ChemHat: Hazards a	an 🛞 ToxCast Das	shboard 2 Q Regula	ations.gov						
ChEMBL	LactMed	C RSC F	Publications						
Consumer Product In	nf International	I Toxicity E & Spring	er Materials						
C ECHA Brief Profile	C ACTOR PDF	Report 🛞 IRIS A	ssessments						
C ECHA Infocard		Q CORE	Literature Sear						
Q Sigma-Aldrich Chemi	c	Q Bielefe	ld Academic Se						

Consuming Available Toxicity Value Data



hemical Properties	Env. Fate/Transport	Synonyms	External Links	Toxicity Values (Beta) E	posure	Bioassays	Similar Mole	ecules (Beta)	Literatur	e Comm	ents
Bioavailability Metric	Download a	as: TSV	Excel									
Exposure Limit					_							
Point Of Departure					Regu	latory Tox	icity Value					
Regulatory Toxicity	v Groupin	g Priority	Туре 🕴 Su	ıbtype Value 🔶	Units 🔶	Study Type 🔶	Exposure Route	Study Duration	Species	Media	Details	Source
Exposure Descriptor	49234	2	RfDo -	0.05	mg/kg-d	-	-	-	-	-	RSL de	RSL
Effect Level	61404	5	RfD -	0.05	mg/kg-d	-	oral	chronic	-	-	RSEI d	RSEI
Misc Hazard Inform	An estimate (with un				de) of	-	oral	-	-	-	EPA Ris	ACToR
	a daily oral exposure subgroups) that is like					-	oral	-	-	-	EPA Ris	ACToR
Screening Level	effects during a lifet BMD, with UFs gene	ime. It can be	e derived from a	NOAEL, LOAEL, o	r i	-	oral	-	-	-	Pennsyl	ACToR
Uncertainty Factor	Generally used in El				J	-	oral	-	-	-	Pennsyl	ACToR
	253004	5	RfD inh	alation 0.0	mg/kg-d	-	inhalation	-	-	-	Pennsyl	ACToR
	253005	5	unit risk inh	alation 0.0	(g/m3)-1	-	inhalation	-	-	-	Pennsyl	ACToR
	253020	4	RfD ora	al 0.05	mg/kg-d	-	oral	_	_	_	Detailed	ACToR





hemical Properties	Env. Fate/Transport	Synonyms	External Links	Toxicity ∀alues (Beta)	Exposure	Bioassays	Similar Molecules (Beta)	Literature	Comments	
Product & Use Cate		as: TSV	Excel	Product	& Use Cate	gories (PUC	5) 🕄			
Chemical Functional		t or Use Categor	rization	Categorization	n type		Number of Uniqu	e Products		•
Monitoring Data	adhesiv	e		CPCat Cassette	e		17			_
Exposure Predictions	manufa	turing		CPCat Cassette	e		17			
	paint			CPCat Cassette	e		16			
	manufa	-		CPCat Cassette	e		12			
	manufa	turing		CPCat Cassette	e		11			
	building flooring	_material		CPCat Cassette	e		8			
	constru	tion		CPCat Cassette	e		8			

Comments

Consumer Product data



 Data gathered from multiple online data sources – text mining and data downloads



- CASRN and name mappings to produce curated structure set with functional uses
- Produce Functional Use (FUSEDB) and build predicted functional use models







National Health and Nutrition Examination Survey

High-Throughput Models for Exposure-Based Chemical Prioritization in the ExpoCast Project

Environ. Sci. Technol., 2013, 47 (15), pp 8479-8488



Product & Use Categori	Download as: TS	V Excel		
Chemical Weight Fraction		National Health and	d Nutrition Examination Survey (NHANES) Infe	rences (ma/ka-bw/day)
Chemical Functional Use		Lower 95th Limit	Upper 95th Limit	Median
Monitoring Data	Ages 6-11	3.80e-05	4.92e-05	4.33e-05
	Ages 12-19	2.55e-05	3.38e-05	2.93e-05
Exposure Predictions	Ages 20-65	2.79e-05	3.27e-05	3.02e-05
	Ages 65+	1.91e-05	2.31e-05	2.10e-05
	BMI > 30	2.38e-05	2.74e-05	2.55e-05
	BMI < 30	3.02e-05	3.30e-05	3.16e-05
	Repro. Age Females	2.83e-05	3.31e-05	3.06e-05
	Females	2.58e-05	3.03e-05	2.80e-05
	Males	2.94e-05	3.37e-05	3.15e-05
	Total	2.86e-05	3.08e-05	2.97e-05

Chemical Properties

External Links Toxicity Values (Beta)

(Beta) Exposure

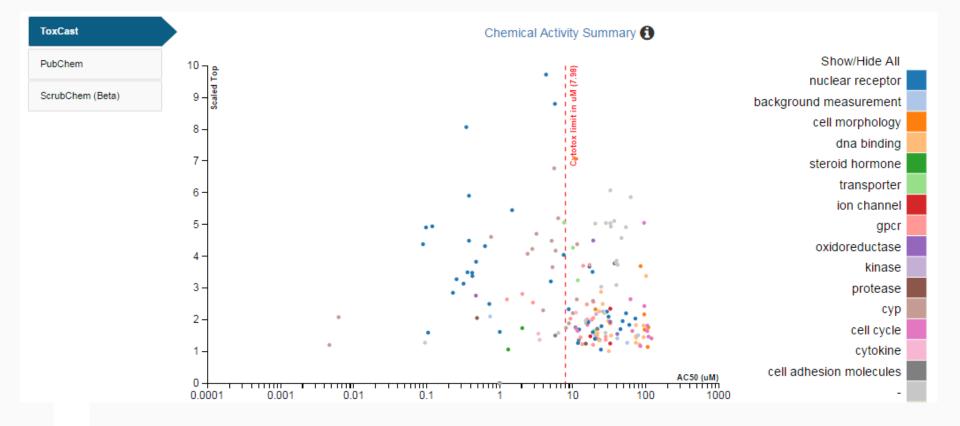
Bioassays

Similar Molecules (Beta)

Literature C

Comments

ToxCast and Tox21 Bioassays



Env. Fate/Transport

Synonyms

External Links

invironmental Protection

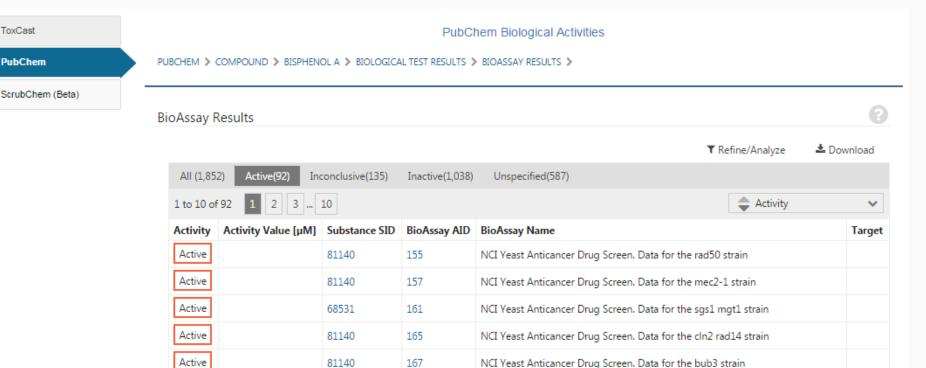
Agency

ToxCast and Tox21 Bioassays



Download as: TSV Excel Show	v: Inactive	Background				
Assay Name	Hit Call	Тор	Scaled Top	AC50	log AC50↓	Intended Target Family
APR_Hepat_CellLoss_48hr_dn	ACTIVE	1.41	1.41	120	2.08	cell cycle
APR_HepG2_MitoMass_24h_dn	ACTIVE	0.874	1.76	109	2.04	cell morphology
APR_HepG2_OxidativeStress_24h_up	ACTIVE	1.20	1.47	110	2.04	cell cycle
APR_Hepat_DNADamage_48hr_up	ACTIVE	1.84	1.14	107	2.03	cell morphology
APR_HepG2_CellLoss_24h_dn	ACTIVE	1.20	1.81	106	2.02	cell cycle
APR_HepG2_OxidativeStress_72h_up	ACTIVE	1.80	1.65	106	2.02	cell cycle
ATG_HSE_CIS_up	ACTIVE	1.59	3.38	102	2.01	dna binding

PubChem Bioassay Data Integration



ToxCast



Literature Search Integration Google Scholar

Chemical Properties

Env. Fate/Transport

Synonyms

External Links



hemical Properties	Env. Fate/Transport	Synonyms	External Links	Toxicity Values (Beta)	Exposure	Bioassays	Similar Molecules (Beta)	Literature	Comments	
Google Scholar		Select Term:	Hazard	•	AND	-	v	AND	-	
PubMed Abstract Sift	er	Edit the Query I	Before Querying (3	39 Characters)						
PubChem Articles		"Hazard" AND) "80-05-7" OR "Bisp	phenol A"						
PubChem Patents				Google	"Hazard" AND	"80-05-7" OR	"Bisphenol A"		•	۹
IRIS		Submit		Scholar	About 17,600 resul	ts (0.14 sec)				
				Articles Case law My library Any time Since 2017 Since 2016 Since 2013 Custom range Sort by relevance Sort by date	evaluation and RW Tyl - Environme Abstract Myers et a Laboratory Practice assessment, using Cited by 53 Relate An updated we bisphenol A a Using a M Wright-Walters, C An aquatic hazard (PNEC) below whic exposure to a chem	risk assessm ental health persp II [Environ Health es (GLPs) cannot bisphenol A (BP ed articles All 13 eight of evider nd the derival C Volz, E Talbott, assessment estal h it is assumed th local. An aquatic F ed articles All 7 v	versus guideline-complian nent: bisphenol A as a ca ectives, 2009 - search.proquest. Perspect 117: 309-315 (2009)] a be used as a criterion for selecti A) as a case study. They did not versions Web of Science: 34 nce approach to the aqua tion a new predicted no ef D Davis - Science of the Total Dishes a derived predicted no ef iat aquatic organisms will not suf nazard assessment of the endoc versions Web of Science: 34 C toxicity	ase study com argued that Good ng data for risk discuss the role (s) Cite Save tic hazard asse effect concentration fect concentration fer adverse effects f rine disruptor Bispt	of essment of tion (Pnec) rom	
				 ✓ include patents ✓ include citations 	HC Alexander, DC I species were not of the chronic envir	Dill, LW Smith t required. One fai onmental hazard urface waters. A t	Environmental, 1988 - Wiley ctor that must be considered in the of bisphenol A is its reportedly piochemical oxygen demand test	he assessment rapid		

Toxicity Values (Beta)

Exposure

Bioassays

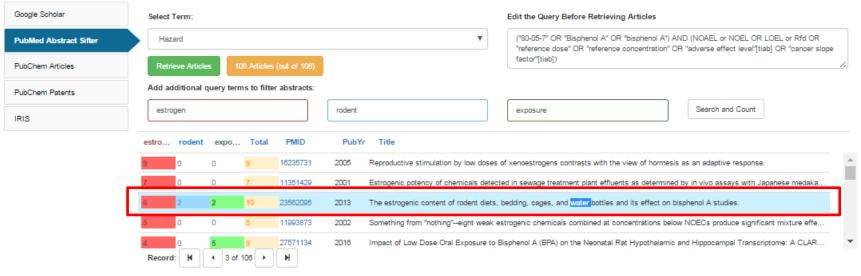
Similar Molecules (Beta)

Literature

Comments

PubMed: BIG DATA Literature Search Integration





Title: The estrogenic content of rodent diets, bedding, cages, and water bottles and its effect on bisphenol A studies.

Abstract: The lowest observed adverse effect level for bisphenol A (BPA) in mice and rats is currently poorly defined due to inconsistent study designs and results in published studies. The objectives of the current study were to (1) compare the **estrogen**ic content of **rodent** diets, bedding, cages, and water bottles to evaluate their impact on the **estrogen**ic activity of BPA and (2) review the literature on BPA to determine the most frequently reported diets, beddings, cages, and water bottles used in animal studies. Our literature review indicated that low-dose BPA animal studies have inconsistent results and that factors contributing to this inconsistency are the uses of high-phytoestrogen diets and the different routes of **exposure**. In 44% (76 of 172) of all reports, **rodent** were exposed to BPA via the subcutaneous route. Our literature review further indicated that the type of diet, bedding, caging, and water bottles used in aDPA studies were not always reported. Only 37% (64 of 172) of the reports described the diet used. In light of these findings, we revoluating low levels of phytoestrogen (less than 20 µg/g diet) and metabolizable energy (approximately 3.1 kcal/g diet) and **estrogen**-free bedding, cages, and water bottles for studies evaluating the **estrogen** cativity of endocrine-disrupting compounds such as BPA. The oral route of BPA **exposure** should be used when results are to be extrapolated to humans.

Chemical Properties

Env. Fate/Transport Synonyms

External Links Toxicity

Toxicity Values (Beta) Exposure

ire Bioassays

s Similar Molecules (Beta)

Comments

Linked Directly to PubMed



PubMed

PubMed comprises more than 27 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

S NCBI Resources 🕑 How To 🕑	tony27587@gmail.com My NCBI Sign Out
Public PubMed VS National Library of Medicine lational institutes of Health Advanced	Search
Format: Abstract - Send to - Am Assoc Lab Anim Sci. 2013 Mar;52(2):130-41. The estrogenic content of rodent diets, bedding, cages, and water bottles and its effect on bisphenol A studies.	Full text links
higpen JE ¹ , Setchell KD, Kissling GE, Locklear J, Caviness GF, Whiteside T, Belcher SM. Brown NM. Collins BJ, Lih FB, Tomer KB, Padilla-Banks E, amacho L, Adsit FG, Grant M.) Author information	Save items
bstract he lowest observed adverse effect level for bisphenol A (BPA) in mice and rats is currently poorly defined due to inconsistent study designs and sults in published studies. The objectives of the current study were to (1) compare the estrogenic content of rodent diets, bedding, cages, and ater bottles to evaluate their impact on the estrogenic activity of BPA and (2) review the literature on BPA to determine the most frequently	Similar articles The effect on sperm production in adult Sprague- Dawley rats exposed by gavag [Toxicol Sci. 2003]
ported diets, beddings, cages, and water bottles used in animal studies. Our literature review indicated that low-dose BPA animal studies have consistent results and that factors contributing to this inconsistency are the uses of high-phytoestrogen diets and the different routes of exposure.	High-fat diet aggravates glucose homeostasis disorder caused by chronic e [J Endocrinol. 2014]
44% (76 of 172) of all reports, rodents were exposed to BPA via the subcutaneous route. Our literature review further indicated that the type of et, bedding, caging, and water bottles used in BPA studies were not always reported. Only 37% (64 of 172) of the reports described the diet	Organizational effects of perinatal exposure to bisphenol-A and diethylstilt [Endocrinology. 2013]
xed. In light of these findings, we recommend the use of a diet containing low levels of phytoestrogen (less than 20 μg/g diet) and metabolizable nergy (approximately 3.1 kcal/g diet) and estrogen-free bedding, cages, and water bottles for studies evaluating the estrogenic activity of ndocrine-disrupting compounds such as BPA. The oral route of BPA exposure should be used when results are to be extrapolated to humans.	Review BPA, an energy balance disruptor. [Crit Rev Food Sci Nutr. 2015]
IID: 23562095 PMCID: <u>PMC3624780</u>	Review Preimplantation Exposure to Bisphenol A and Triclosan May Leac [Biomed Res Int. 2015]
ndexed for MEDLINE) Free PMC Article	See reviews See all

Answering Questions



- I have a 1000 CAS Numbers (or Names) is there data available?
 - Has any Toxcast data been run?
 - Are there Toxicity Data values available?
 - Are there predicted exposure data (via Expocast)?
 - Can I get predicted physchem data for my model?
- Identifying chemicals by analytical chemistry

Mass Spectrometry Analyses

• Targeted Analysis:

- We know exactly what we're looking for
- 10s 100s of chemicals

• Suspect Screening Analysis (SSA):

- We have chemicals of interest
- 100s 1,000s of chemicals
- Non-Targeted Analysis (NTA):
 - We have no preconceived lists
 - 1,000s 10,000s of chemicals
 - In dust, soil, food, air, water, products, etc.

NTA Applications





- Exposure surveillance
 - What chemicals are in food, products, dust, blood, etc.?
- Chemical prioritization
 - What are relevant chemicals & mixtures?
- MBRITA
- Exposure forensics
 - What are chemical signatures of exposure sources?



- Effect-directed analysis
 - What are the biologically active chemicals in complex mixtures?
- Biomarker discovery
 - What chemicals are predictive of bioactivity/health impairment?

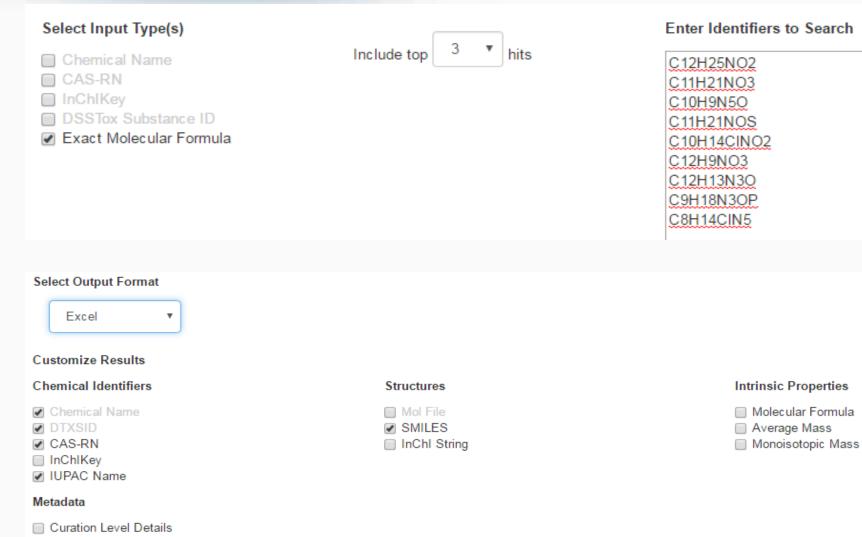






Batch Searching - Formulae





Data Sources

Batch Searching - Formulae

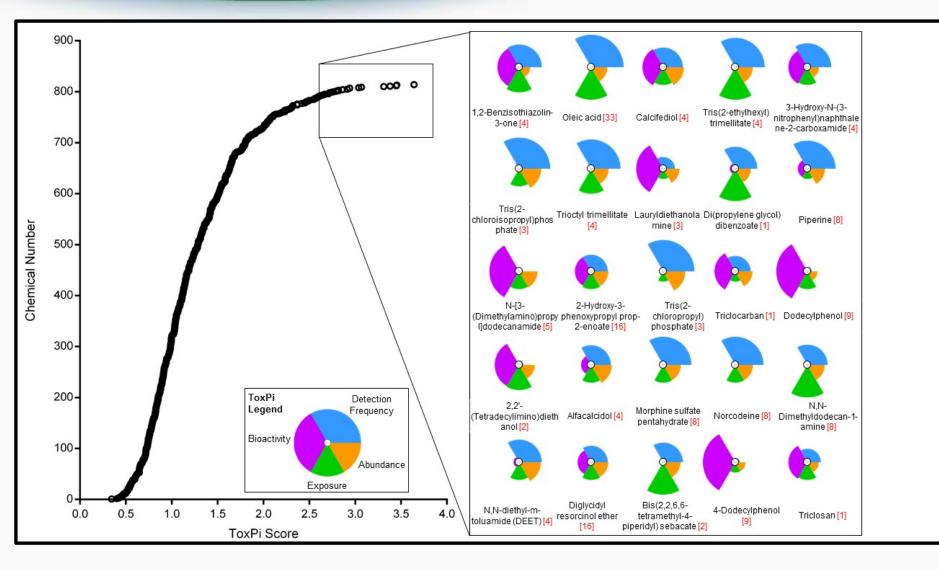


- Uses data source ranking
- Production volume ranking in development

	A	В	С	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3		19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23CIN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S		35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35CIN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35CIN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O		40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13CIN2O		11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxypyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026		N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3		68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3		51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19CIN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2,		290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quina	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-		290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxypheny	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3

Chemical Prioritization: Dust Case Study





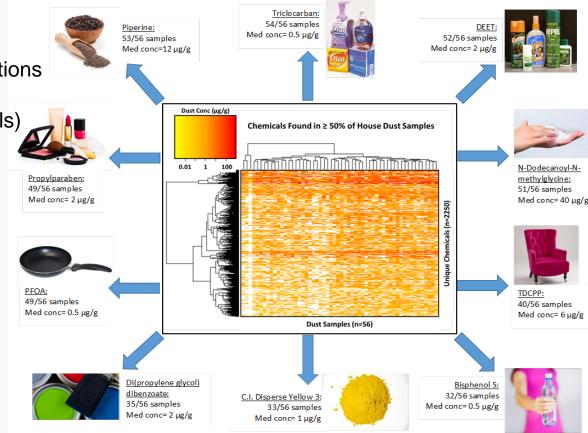
Rager et al., *Environ. Int.*, v88, March 2016, pp 269–280.

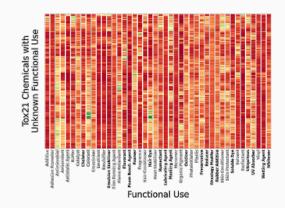
DOI: 10.1016/j.envint.2015.12.008

Suspect Screening Results: Dust Case Study



- Chemical confirmation from ToxCast
- Detection frequency
- External calibration for concentrations
- Cluster analysis (homes/chemicals)
- Chemical functional use





Improving Exposure Estimates and Characterization



100 Consumer Products and Articles of Commerce

- GCXGC-MS with DCM Extraction
- 1606 tentative and confirmed chemical identifications
- Common Chemical (n>19)
- ToxCast
- Flame Retardant
- Potent ER

- Log₁₀ (μg/g) Air freshener Baby soap Carpet
 - Carpet padding
 - Cereals
 - Cotton clothing
 - Deodorant
 - Fabric upholstery
 - Glass cleaners
 - Hand soap
 - Indoor house paint
 - Lipstick
 - Plastic children's toys
 - Shampoo
 - Shaving cream
 - Shower curtain
 - Skin lotion
 - Sunscreen
 - Toothpaste
 - Vinyl upholstery

Wambaugh et al. Unpublished

"RapidTox" in Development



- Semi-automated decision support tool for highthroughput risk assessments
- Use the Dashboard "architecture" and data streams and add appropriate functionality
- Combine diverse data streams into quantitative toxicity values with uncertainty estimates
- Test out new prototype capabilities for internal users – e.g. GenRA (Generalized Read Across)

Risk Assessments Generally Contain a Standard Set of Components



\$epa	United States Environmental Protection Agency	EPA Document# 740-R1-5002 March 2015 Office of Chemical Safety and Pollution Prevention
TSC/	A Work Plan Chemical Risk	Assessment
	N-Methylpyrrolidon Paint Stripper Use	
	CASRN: 872-50-4	
	N N N	
	March 2015	

TABLE OF CONTENTS	
TABLE OF CONTENTS	. 2
AUTHORS / CONTRIBUTORS / ACKNOWLEDGEMENTS / REVIEWERS	. 9
ABBREVIATIONS	11
EXECUTIVE SUMMARY	
1 BACKGROUND AND SCOPE	20
1.1 INTRODUCTION	
1.2 USES AND PRODUCTION VOLUMES 1.2.1 Assessment and Regulatory History	
1.2.1 Assessment and Regulatory History	
1.3 PROBLEM FORMULATION	23
1.3.1 Physical ar Corr leal Reoperties	24
13.1 Physical ar Drivel Boord's Chem	25
1.3.3 Conceptual Model	26
1.3.3.1 Exposure Pathways 1.3.3.2 Health Effects and Human Recentors	
1.3.2 Health Ellects and Human Receptors.	
2 EXPOSURE ASSESSMENT	
2.1 OCCUPATIONAL EXPOSURES	
2.1.1 Approach and Methodology 2.1.1 Identification of Relevant Industries	
2.1.1.2 Approach for Determining Occupational Exposure Data and Input Parameters for PBPK Modelin	g
2.1.1.3 Estimates of Occupational Exposure Parameters and Number of Exposed Workers.	
2.1.1.3 Estimate conficupational Exposure Parameters and Number of Exposed Workers	
	37
2.2.1 Approach and Methodology	
2.2.1.1 Consumer Dermal Exposure Assessment	
2.2.1.2 Consumer Users and Residential Non-Users Inhalation Exposure Assessment	
2.2.2 Model Outputs and Exposure Calculations	
3 HAZARD IDENTIFICATION AND DOSE-RESPONSE	48
3.1 APPROACH AND METHODOLOGY	
3.1.1 Selection of Peer-Reviewed Assessments for Hazard Identification and Dose-Response Analysis	
3.1.2 Hazard Summary and H zar Identification 3.1.3 Selection of Developmentary wards St. die 2. d. ind for its	
3.1.3 Selection of Development on the Static and Indian Structure 3.1.3.1 Decreased Fetal and Faster theory of the Static and Structure 3.1.3.2 Resortions and Fetal Mortality	63
3.1.3.2 Resorptions and Fetal Mortality	65
3.1.3.3 Other Fetal Effects	
3.1.3.4 Conclusions and Selection of Key Endpoints	
3.2 DOSE-RESPONSE ASSESSMENT AND STOLD SELECTION.	68
DAGA of Pareno neo DK 31	ä.
Dose Kesponse, PK, al	73
3.2.4 PODs for Chronic Exposure	76
PODS	

3.2.5	Considerations for Sensitive Subpopulations and Lifestages
и ними	
HUM/	AN HEALTH RISK CHARACTERIZATION
4.1 RISK ES	
4.1.1	Rick Eximus : A loute Oco p tip a Exios in to IN P 🕤
4.1.2	Risk Estimates for Acute Consumer Exposure to NMP
4.1.1	Risk Estimates for Chronic Occupational Exposures to NMP
	N HEALTH RISK CHARACTERIZATION SUMMARY
	URCES OF UNCERTAINTY AND DATA LIMITATIONS
4.3.1	Key Uncertainties in the Occupational Exposure Assessment
4.3.2	Key Unce tain le linch Con mar kpesures sessi en
4.3.3	Key Uncernettier in heterization Done here as the second
4.3.4	Key Uncertainties in the Risk Assessment
4.4 RISK AS	SESSMENT CONCLUSIONS
REFERENCES	
PPENDICES	
Appendix A	ENVIRONMENTAL EFFECTS SUMMARY
	OXICITY TO AQUATIC ORGANISMS
	TOXICITY TO AQUATIC ORGANISMS
	TO WILDLIFE
A-5 SUMMA	
Appendix B	CHEMICAL REPORTING DATA 12
B-1 CONSUL	IER USES 12
	RIPPING APPLICATIONS 12
Appendix C	STATE NMP REGULATIONS 12
Appendix D	OCCUPATIONAL EXPOSURE ASSESSMENT SUPPORT INFORMATION
	RY OF DERMAL EXPOSURE PARAMETERS, INHALATION CONCENTRATIONS AND EXPOSURE REDUCTION FACTORS13
	EEDS AND DATA COLLECTION
	IES THAT EMPLOY PAINT STRIPPING ACTIVITIES
	TIONAL PAINT STRIPPING PROCESSES AND ASSOCIATED WORKER ACTIVITIES
D-4 Occupa	AND POPULATION DATA AND INFORMATION
D-4 Occupa D-5 Facility	
D-4 Occupa D-5 Facility D-6 Dermai	Exposure Parameters
D-4 Occupa D-5 Facility D-6 Dermai	EXPOSURE MARAMETERS
D-4 Occupa D-5 Facility D-6 Dermai	
D-4 Occupa D-5 Faciliti D-6 DERMAI D-7 Occupa Appendix E	TIONAL IMMALATION EXPOSURE LITERATURE DATA
D-4 Occupa D-5 Faciliti D-6 Dermai D-7 Occupa Appendix E E-1 Estimat	INNAL INHALATION EXPOSURE LITERATURE DATA
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa Appendix E E-1 Estimat E-2 Sensitiv	TIONAL INNALATION EXPOSURE LITERATURE DATA
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa Appendix E E-1 Estimat E-2 Sensity E-3 Inhalat	NONALITION EXPOSURE LITERATURE DATA 14 CONSUMER EXPOSURE ASSESSMENT 15 IONI OF ENDISON PROFILES FOR PAINT FERMOVERS/STREPPERS 15 ITT ANALYSIS FOR INALATION SCHARIOS 16 INFERIORIES FOR INALATION SCHARIOS 16
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa Appendix E E-1 Estimat E-2 Sensiti E-3 Innalat E-4 Innalat	Instal Instalation Exposure Litreature Data 14 CONSUMER EXPOSURE ASSESSMENT 15 Ioin of Exission Propriets for Paint Removers/Strepress 15 Ioin of Exission Propriets for Paint Removers/Strepress 16 In Adults for bill match on Scenarios 16 Ion Evision Propriets Stream Instrume 16 Ion Evision Propriets Stream Instrume 17 Model Output Stream Instrume 17
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa Appendix E E-1 Estimat E-2 Sensiti E-3 Innalat E-4 Innalat	Instalation Exposure Literature Data 14 CONSUMER EXPOSURE ASSESSMENT 15 Iono of EMISSION PROFILES FOR PAINT (ERMOVERS/STREPPERS 15 Instruction Section and Constraints 16 Ion Model Outputs and Endoversity 18
D-4 Occupa D-5 Facility D-6 DERMAN D-7 Occupa D-7 Occupa E-1 Estimat E-2 Sensity E-3 Inhalat E-4 Inhalat E-5 MCCEN	Instal Instalation Exposure Litreature Data 14 CONSUMER EXPOSURE ASSESSMENT 15 Ioin of Exission Propriets for Paint Removers/Strepress 15 Ioin of Exission Propriets for Paint Removers/Strepress 16 In Adults for bill match on Scenarios 16 Ion Evision Propriets Stream Instrume 16 Ion Evision Propriets Stream Instrume 17 Model Output Stream Instrume 17
D-4 Occupa D-5 Facility D-6 DERMAI D-7 Occupa D-7 Occupa E-1 Estimat E-2 SENSITI E-3 INHALAT E-4 INHALAT E-5 MCCEN	Instalation Exposure Litreature Data 14 CONSUMER EXPOSURE ASSESSMENT 15 Isson of Exposure Assessment 15 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 17 Instalation Science Science 17 Instalation Science Case Summers 16 INMP Science Into Lington Ethics, Brush-On, Workshop, User In ROH during wait time, 0.45 ACH, 0.2 18 INMP Sciencia 16
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa oppendix E E-1 Estimat E-2 Sensity E-3 Inhalat E-4 Inhalat E-5 MCCEN <i>E-5-1</i>	Instal Instalation Evrosule Litreature Data 14 CONSUMER EXPOSUME ASSESSMENT 15 ion or Eurosion Propries ron Pauri Removes/Strappers 15 ion or Eurosion Propries ron Pauri Removes/Strappers 16 ion Eurosion Propries ron Pauri Removes/Strappers 16 ion Eurosion Propries Constrainting Constraintin Constrainting Constrainting Constrainting Constrainting Constra
D-4 Occupa D-5 Facility D-6 Dermai D-7 Occupa Appendix E E-1 Estimat E-2 Sensity E-3 Inhalat E-4 Inhalat E-5 MCCEN <i>E-5-1</i>	Instalation Exposure Litreature Data 14 CONSUMER EXPOSURE ASSESSMENT 15 Isson of Exposure Assessment 15 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 16 Instalation Science Science 17 Instalation Science Science 17 Instalation Science Case Summers 16 INMP Science Into Lington Ethics, Brush-On, Workshop, User In ROH during wait time, 0.45 ACH, 0.2 18 INMP Sciencia 16

We are assembling these components to deliver RapidTox

Problem #1 – SuperFund site





- At a SuperFund site, EPA has identified 600 unique chemicals
- 300 of these have "good" reference doses/concentration (RfD/RfC), but 300 do not



- Can we determine "good enough" RfD/RfC to aid in cleanup planning?
- Can we do it "quickly"?

Problem #2 – Prioritize inerts



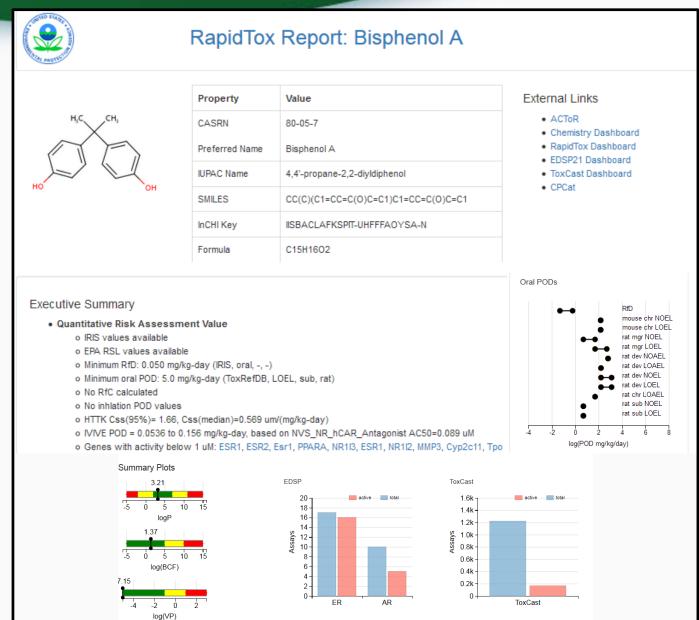


LAWN

- Office of Pesticide Programs (OPP) petitioned to perform risk assessments on 100s of pesticidal "inert ingredients"
- Companies have **not** typically been required to submit *in vivo* data on individual inerts
- Can we **prioritize** which of these chemicals should be the focus of detailed risk assessments?

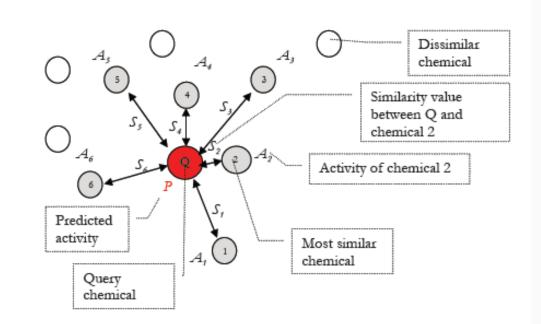
RapidTox Dashboard in development Risk Assessment Tool





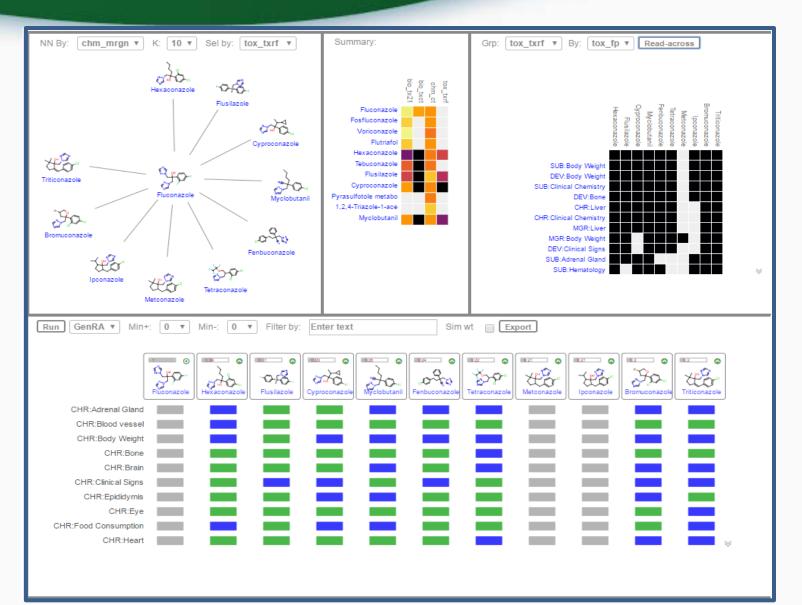


- "Chemical-Biological Read-Across" predict toxicity as a similarity-weighted activity of nearest neighbors
- Evaluates read-across performance and uncertainty using available data



Work-in-progress





Deliver Data and Models for Reuse



Toxicity ForeCaster (ToxCast[™]) Data

EPA's most updated, publicly available high-throughput toxicity data on thousands of chemicals. This data is generated through the EPA's ToxCast research effort. ToxCast is part of the Toxicology in the 21st Century (Tox21) federal collaboration. All data is available for download and includes the following data sets. The release date and version names for the data sets are provided in the table below.

As part of EPA's commitment to share data, all of the computational toxicology data is publicly available for anyone to access and use. EPA's computational toxicology data is considered "open data", and thus all of the data below are free of all copyright restrictions, and fully and freely available for both non-commercial and commercial use.



- ToxCast & Tox21 Chemicals: A list of all chemicals screened, along with descriptions.
- ToxCast & Tox21 High-throughput Assays: Information about the hundreds of assays used to screen the chemicals.
- ToxCast & Tox21 Summary Data: Summary hit calls for all chemicals and concentration response curves for all assays.
- MySQL Database: Downloadable database that provides access to EPA's analysis of the ToxCast and Tox21 high-throughput screening data.
- **R Package**: The computer programming package used to process and model all EPA ToxCast and Tox21 chemical screening data.
- Data Spreadsheet: Spreadsheet that provides access to EPA's analysis of the ToxCast and Tox21 high-throughput screening data.
- Concentration Response Curves: Concentration response curves for all ToxCast & Tox21 assays.
- Collaborative Estrogen Receptor Activity Prediction Project Data: Data and supplemental files from CERAPP, a large-scale modeling project.

https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data

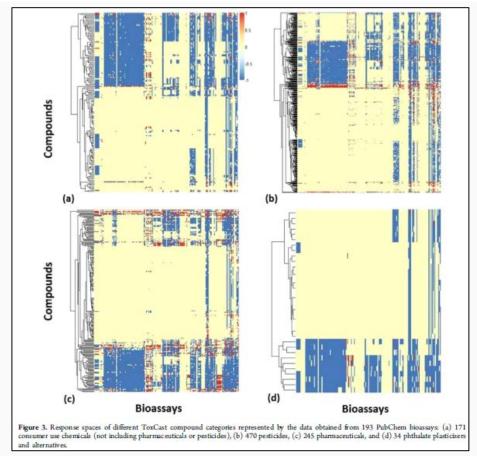
Open Data means Reuse





Big Data in Chemical Toxicity Research: The Use of High-Throughput Screening Assays To Identify Potential Toxicants

Hao Zhu,*^{,†,‡} Jun Zhang,^{†,‡} Marlene T. Kim,^{†,‡} Abena Boison,[†] Alexander Sedykh,[‡] and Kimberlee Moran[§]



Delivering Data to the Community



• Linkages are simple: using the associated identifiers



Searched by Synonym: Found 1 result for 'bisphenol A'.

2D	3D	Q	[dif]	ß	*			
							Intrinsic Properties	
		H ₃ C		,CH	2		Structural Identifiers	
		,	$\left \right>$,		Citation	
			Ś	F			U.S. Environmental Protection Agency. iCSS Chemistry Dashboard. https://comptox.epa.gov/dashboard/DTXSID7020182 (accessed May 03, 2016), Bisphenol A	
но						он		6

- Files of structures/identifiers mapped to DTXSIDs
- Integrated already PubChem, EBI's UNICHEM, ChemSpider, ToxPlanet. Whoever wants the files...

Delivering our Chemistry Data

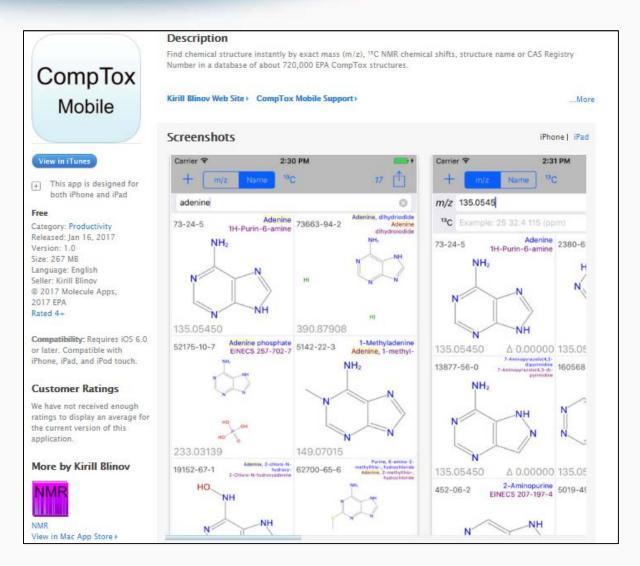


~	Environmen	es ntal Protection	Home	Advanced Search	Batch Search	Lists	Search Chemistry Dashl	board	Q	Op
Ch	emistry	Dashboar	d						Aa 🕶	Aa
					Downlo	ads				
DS	STox Identifier to	PubChem Identifier	Mapping F	ïle					Posted: 11/14/20	016
The	DSSTox to Pub	Chem Identifiers ma	pping file is	in TXT format and incl	udes the PubChem	SID, PubChem Ci	ID and DSSTox substance ident	tifier (DTXSID).		
	SID 316388 316388 316388 316388 316388 316388 316388 316388 316388	890 889 888 887 886 885	10 50 19 11 25	404 142816 742127 073841 505215 021861 84427	DT DT DT DT DT DT DT	XSID XSID308 XSID708 XSID408 XSID208 XSID008 XSID808 XSID608 XSID608	73142 73139 73137 73135 73133 73133 73131			
DC	CT	napped to CAS Nun		lanaa Tila					Posted: 11/14/20	048
					ar DSSTay substan	oo identifier (DTVS	SID) and the Preferred Name.		r usteu. 11/14/20	010
	A A	B	annat and li	induce the one numb	a, boorox substan	ou wentiner (DTA:	and the merened Name.			
1	casrn	dsstox subst	ance id	preferred name						
2		DTXSID702000		A-alpha-C						
3		DTXSID202000	4	Acetaldehyde oxime						
4	60-35-5	DTXSID702000	5	Acetamide						
5	103-90-2	DTXSID202000	6	Acetaminophen						
6		DTXSID702000	7	Acetohexamide						
		DTXSID202000		Acetone[4-(5-nitro-	2-furyl)-2-thiazoly] hydrazone				
7		DTXSID702000	0	Acetonitrile						
8										
8 9	75-05-8 127-06-0 65734-38-5	DTXSID602001 DTXSID602001	0	Acetoxime N'-Acetyl-4-(hydrox)		0				

 Various types of data at FTP download site: <u>ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_</u> <u>Data/Chemistry_Dashboard</u>
 62

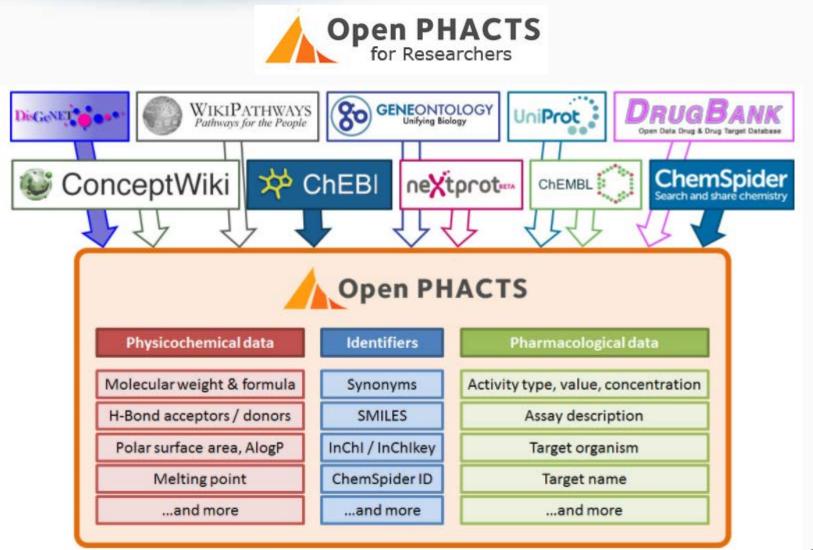
Open Data Reuse on iOS





Delivering More Value from Our Data – TRIPLE STORE?





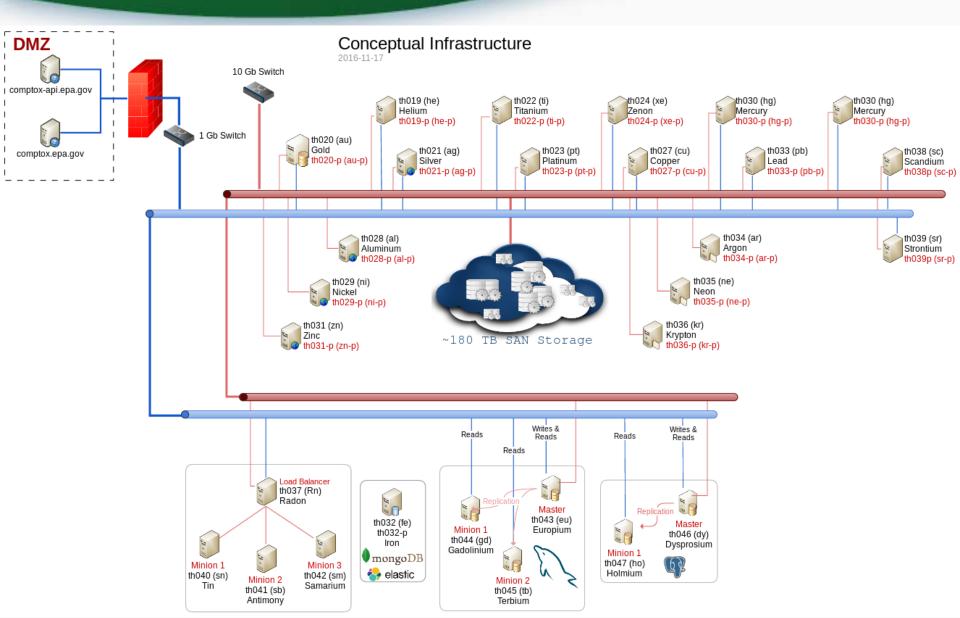
Computational Infrastructure and Processes



- 10 years of development heterogeneous
 - MySQL, Java, R-code, Matlab, multiple Javascript environments, 100s of web services
 - Multiple projects, loose connectivity
- Migration to homogeneity for production Apps
 - MySQL
 - Java
 - JQuery/JavaScript
 - Versioned microservices
- Sprint methodology with Atlassian management tools
 - **Confluence** knowledge management
 - Jira ticketing/stories (work assignments)
 - Fisheye collaborative peer code review
 - Bitbucket code version control system

Computational Infrastructure





The NCCT Team

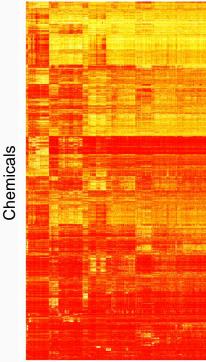


- Our team is broad in skills
 - Biologists, chemists and toxicologists
 - Bioinformaticians and cheminformaticians modelers (QSAR, Deep Learning), text miners
 - Information technology team, software developers
 - A large IT support team for production applications (National Computing Center)

Many Successes from **Our Approaches**



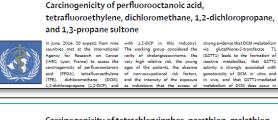
Group Chemicals by Similar Bioactivity and **Predictive Modeling**



Assays/Pathways

Provide **Mechanistic** Support for Hazard ID

Prioritization of Chemicals for Further Testing



Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate



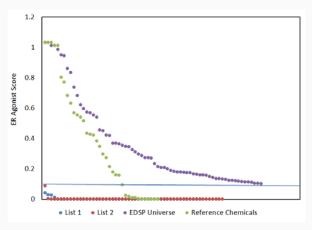
herbicide 2.4-dichlorophenoxyacetic. It was widely applied to eradicate

March, 2015, 17 experts from cell proliferation (hyperplasia in The insecticides malathion 11 countries met at the International rodents). Tetrachlorvinphos is banned diazinon were classified as "probably Agency for Research on Cancer (IARC: in the European Union. In the USA. carcinogenic to humans" (Group 2A) on, France) to assess the carcino- it continues to be used on animals, Malathion is used in agricult ure, public genicity of the organophosphate including in pet flea collars. pesticides tetrachlorvinphos, parathion, For parathion, associations with health and residential insect control It continues to be produced alathion, diazinon, and glyphosate cancers in several tissues were substantial volumes throughout the table). These assessments will be observed in occupational studies, world. There is limited ev

Carcinogenicity of lindane, DDT, and 2,4-dichlorophenoxyacetic acid In June, 2015, 26 experts from Immunosuppressive effects that can blood or adipose taken in adulthood; 13 countries met at the International operate in humans. however, the possible importance of Agency for Research on Cancer The Insecticide DDT was classified early-life exposure to DDT remains (IARC; Lyon, France) to assess the as "probably carcinogenic to humans" unresolved. Studies on non-Hodgkin arcinogenicity of the insecticides (Group 2A). DDT was used for the lymphoma and cancers of the liver Indane and 1.11-trichloro-2.2-bis(4- control of Insect-borne diseases and testis provided limited evidence chlorophenyl)ethane (DDT), and the during World War 2; subsequently In humans

IARC Monographs 110, 112, 113

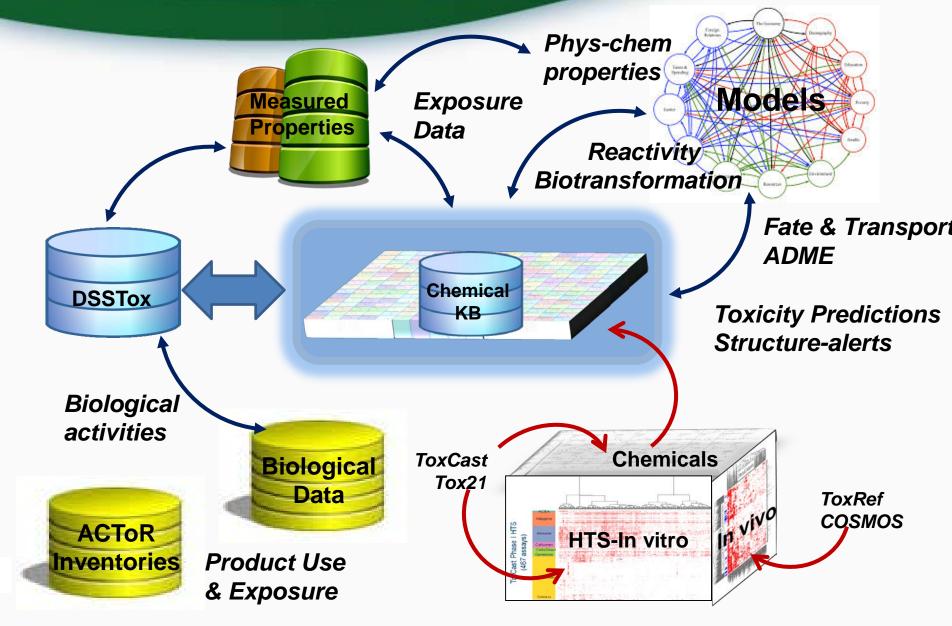
of DDT



FIFRA SAP, Dec 2014

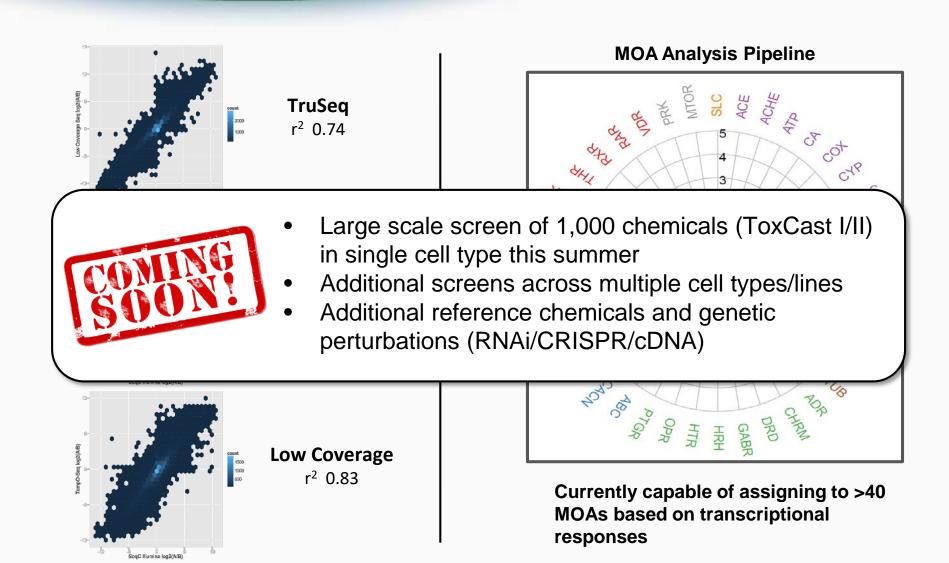
We're not done yet...





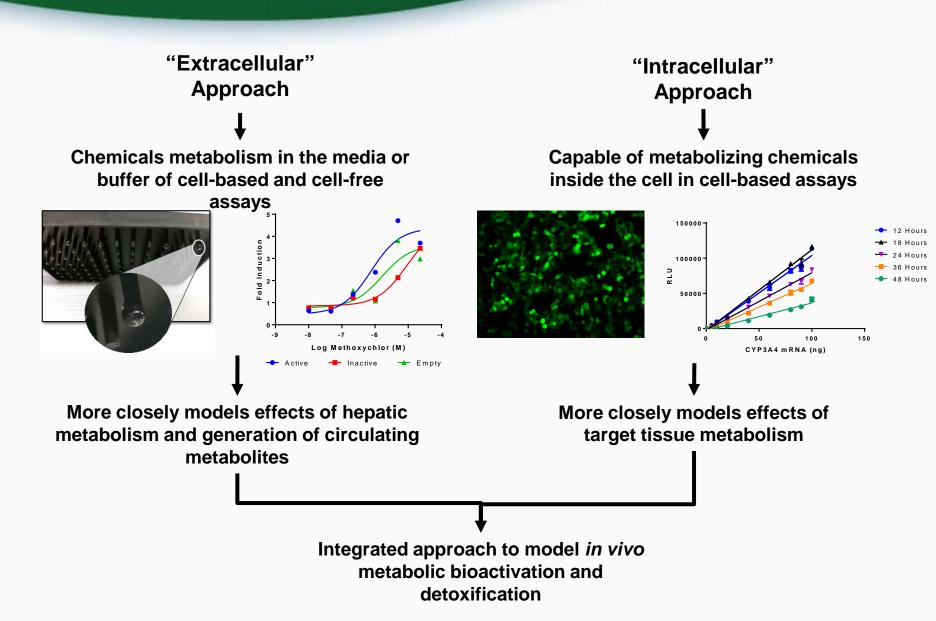
Transcriptomics Data will Deliver Terabytes of Data for Analysis





Beginning to Address Metabolic Competence





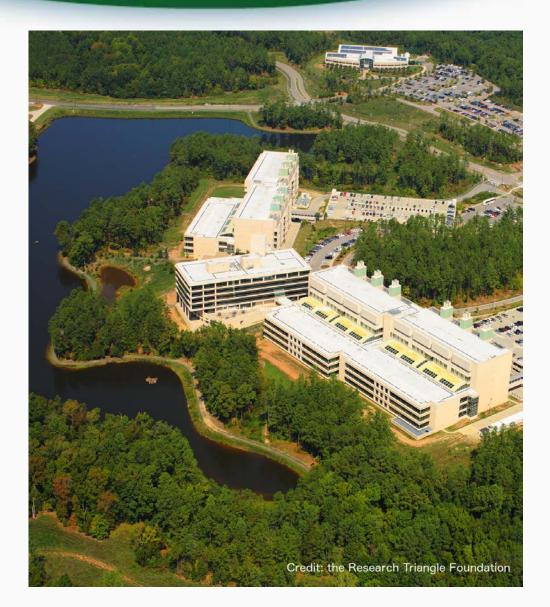
Future Work



- Existing dashboards functionality to be added to the "CompTox Chemistry Dashboard"
- RapidTox functionality tested and released
- Open Application Programming Interface and web services
- Integration to other agency databases
- Deliver our data to the semantic web
 - Use accepted ontologies, chemical structures first

Acknowledgements





EPA-RTP Russell Thomas Kevin Crofton John Cowden Chris Grulke Ann Richard John Wambaugh Grace Patlewicz Imran Shah Richard Judson Jeff Edwards Nancy Baker NERL Colleagues



Antony Williams

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

Williams.Antony@epa.gov

ORCID: https://orcid.org/0000-0002-2668-4821

imental Protection