

The EPA Comptox Chemistry Dashboard NCCT: Antony Williams, Chris Grulke, Jennifer Smith, Kamel Mansouri, Andrew McEachran, Ann Richard and Jeff Edwards

Science Context

A consolidated web platform is necessary for researchers to access chemical information look-up, models and model predictions and linkages to Agency and public resources. This will provide access to: curated chemical structures, computed and measured physchem properties, exposure and usage patterns, data analysis workflows, reaction chemistry, and Quantitative Structure-Activity Relationship (QSAR)-based physiologically based pharmacokinetic (PBPK) and toxicity predictions. The web platform will be extended to include additional capabilities to support "read-across" approaches. Such resources will be made available for broadest possible usage by cheminformaticians, modelers, non-chemists and risk assessors.

The views expressed in this poster are those of the author[s] and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Approach

1) Develop underlying componentized architecture utilizing a "build once, use many times" approach



- 2) Deliver web application access to DSSTox database (ca. 720,000 structures). Beta version delivered April 2016 at ACS Spring.
- 3) Integrate other data via web services (CPCat, Toxcast, ExpoCast data) for Version 1 release, August 2016 at ACS Fall Meeting.
- 4) Utilize Chemistry Dashboard as underpinning architecture to develop RapidTox prototype by August 2016
- 5) Demonstrate value of dashboard to support "Non-Targeted Analysis" to examine the Exposome (*publication submitted*)





epa.gov/research

NERL: Jon Sobus, Kathie Dionisio and Katherine Phillips

NRMRL: Todd Martin

Preliminary Results

				Search			ch CompTox Dashboard	
						Sub	mit Comment	Sh
esult for 'bispher	nol A'.							
Int	rinsic Properties							
1	Molecular Form	nula: C15H1	1602			Q, Find	All Chemicals	
	Average Mass:	228.291 g/n	nol					
1	Monoisotopic I	Mass: 228.1	15030 g/mol				ß	
Str	uctural Identifier	c						
Str	uctural Identifier	s						
Str	uctural Identifier cord Information	s						
Str	uctural Identifier	s						
onyms Product	uctural Identifier cord Information Composition	s I ToxCast in Vitro	o Data Exp	osure PubC	hem Con	nments		
onyms Product	uctural Identifier cord Information Composition	s I ToxCast in Vitro	o Data Exp	osure PubC	hem Con	nments		
nonyms Product s: CSV Excel	uctural Identifier cord Information Composition	s I ToxCast in Vitro	o Data Exp	oosure PubC	hem Con	nments		
nonyms Product	uctural Identifier cord Information Composition	s ToxCast in Vitro Median	Data Exp Range	oosure PubC	hem Con	nments		
onyms Product	uctural Identifier cord Information Composition SDF Average (Exp.)	s ToxCast in Vitro Median (Exp.)	Range (Exp.)	Average (Pred.)	hem Com Median (Pred.)	nments Range (Pred	.) Result U	nit
r Partition ogP)	Composition SDF Average (Exp.) 3.38 (2)	S ToxCast in Vitro Median (Exp.) 3.43	Range (Exp.) 3.43	Average (Pred.) 3.42 (2)	hem Con Median (Pred.) 3.42	Range (Pred 3.20 to 3.64	.) Result U	nit
str Re CSV Excel rr Partition ogP)	Composition SDF Average (Exp.) 3.38 (2) 5.26e-04 (1	s ToxCast in Vitro Median (Exp.) 3.43) 5.26e-04	Data Exp Range (Exp.) 3.43 5.26e-04 5.26e-04	Average (Pred.) 3.42 (2) 2.22e-03 (2)	hem Con Median (Pred.) 3.42 2.22e-03	Range (Pred 3.20 to 3.64 7.56e-04 to	.) Result Ur - mol/L	nit
str Re nonyms Product CSV Excel er Partition .ogP) lity	cord Information Composition Composition C	Median (Exp.) 3.43) 5.26e-04	Data Exp Range (Exp.) 3.43 5.26e-04 163 to 168	Average (Pred.) 3.42 (2) 2.22e-03 (2)	hem Con Median (Pred.) 3.42 2.22e-03	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03	.) Result Un - mol/L	nit
r Partition ogP)	Composition Compositin Composition Composition Composition Composition Composi	s ToxCast in Vitro Median (Exp.) 3.43) 5.26e-04 156 200	Data Exp Range (Exp.) 3.43 5.26e-04 153 to 158 200 200	Average (Pred.) 3.42 (2) 2.22e-03 (2) 4 138 (2) 349 (2)	hem Con Median (Pred.) 3.42 2.22e-03 138 349	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364	.) Result Un - mol/L °C °C	nit
s: CSV Excel	Composition SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1)	s FoxCast in Vitro Median (Exp.) 3.43) 5.26e-04 156 200 -	Data Exp Range (Exp.) 3.43 5.26e-04 153 to 158 200 -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 4 138 (2) 349 (2) 7.06e-08 (1)	hem Con (Pred.) 3.42 2.22e-03 138 349 7.06e-08	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364	.) Result U - mol/L °C °C mmHg	nit
CSV Excel CSV E	Composition SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1) -	s FoxCast in Vitro Median (Exp.) 3.43) 5.26e-04 156 200 - -	Data Exp Range (Exp.) 3.43 5.26e-04 153 to 158 200 - - -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 3.49 (2) 7.06e-08 (1) 2.92 (2)	Median (Pred.) 3.42 2.22e-03 138 349 7.06e-08 2.92	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364 - 2.74 to 3.10	.) Result U - mol/L °C °C mmHg -	nit
er Partition .ogP) lity Partition Coefficient	Composition SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1) - -	s FoxCast in Vitro Median (Exp.) 3.43) 5.26e-04 156 200 - - - -	Data Exp Range (Exp.) 3.43 5.26e-04 153 to 158 200 - - - -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 4 138 (2) 349 (2) 7.06e-08 (1) 2.92 (2) 8.39 (1)	Median (Pred.) 3.42 2.22e-03 138 349 7.06e-08 2.92 8.39	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364 - 2.74 to 3.10 -	.) Result Un - mol/L °C mmHg - -	nit
monyms Product IS: CSV Excel ter Partition LogP) illity it immediate Coefficient Partition Coefficient Hydroxylation Rat	Composition SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1) - - - t - e -	s ToxCast in Vitro Median (Exp.) 3.43 156 200 - - - - - - - -	Data Exp Range (Exp.) 3.43 5.26e-04 153 to 158 200 - - - - - - - - - - - - - - - - - - -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 349 (2) 7.06e-08 (1) 2.92 (2) 8.39 (1) -10.4 (1)	Median (Pred.) 3.42 2.22e-03 138 2.42 3.42 2.33 3.42 2.33 3.43 3.43 3.42 2.33 3.43 3.43 3.43 3.43 3.43 3.43 3.43 3.43 3.43 3.43 3.43	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364 - 2.74 to 3.10 - -	.) Result Un - mol/L °C °C mmHg - - cm3/mole *sec	nit
ronyms Product s: CSV Excel ter Partition LogP) ility t t partition Coefficient Partition Coefficiert Hydroxylation Ral ion Half Life	uctural Identifier cord Information Composition 2 SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1) - - - - -	s FoxCast in Vitro Median (Exp.) 3.43 156 200 - - - - - - - - - - - -	Data Exp Range (Exp.) 3.43 3.43 5.26e-04 153 to 158 200 - - - - - - - - - - - - - - - - - - - - - - - - - - - -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 3.49 (2) 7.06e-08 (1) 2.92 (2) 8.39 (1) -10.4 (1) 15.1 (1)	Median Con Median 2.22e-03 138 2.342 138 2.92 8.39 - -10.4 -	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364 - 2.74 to 3.10 - -	.) Result Un - mol/L °C °C °C mmHg - cm3/mole *sec days	nit
st Re nonyms Product s: CSV Excel er Partition LogP) ility t t rre on Coefficient ² artition Coefficient Hydroxylation Rat on Half Life tion Factor	uctural Identifier cord Information Composition 2 SDF Average (Exp.) 3.38 (2) 5.26e-04 (1) 155 (7) 200 (1) - - - t - - - - - - -	s FoxCast in Vitro Median (Exp.) 3.43 156 200 - - - - - - - - - - - - -	Data Exp Range (Exp.) 3.43 5.266-04 158 to 158 200 153 to 158 200 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	Average (Pred.) 3.42 (2) 2.22e-03 (2) 349 (2) 7.06e-08 (1) 2.92 (2) 8.39 (1) -10.4 (1) 15.1 (1) 173 (1)	Median Con Median	Range (Pred 3.20 to 3.64 7.56e-04 to 3.68e-03 132 to 144 334 to 364 - 2.74 to 3.10 - 2.74 to 3.10 - -	.) Result Un - mol/L °C °C °C mmHg - cm3/mole *sec days	nit

General	loxicology	1 ablications	Analytical	Trediction					
EPA Substance R	E ACToR	Toxline	Q National Environm	Proton NMR Predi					
NIST Chemistry W •	DrugPortal	Environmental He	RSC Analytical Ab	Carbon-13 NMR					
& Household Produ	CCRIS	NIEHS	MONA: MassBan	2D NMR HSQC/	Eve	forn	allin	kc +	
🙄 PubChem 🤅	€) ChemView	National Toxicolog	Tox21 Analytical D	ChemRTP Predictor	EXL	erno	AI LIII	KS L	UE
🕅 Chemspider	CTD	G Google Books							
CPCat 0	👷 eChemPortal	G Google Scholar			ΙΠΤΕ	erag(ency	ana	i pi
🥬 DrugBank 🤅	EDSP Dashboard	G Google Patents				0		-	•
Mage HMDB	Gene-Tox	PubMed				L	vebs	ites	
w Wikipedia	HSDB	Q BioCaddie DataMed				-			
Q MSDS Lookup	ToxCast Dashboar								
Q ToxPlanet	LactMed								
Q ChemHat: Hazard	International Toxici								
Consumer Produc	ACTOR PDF Report								
				Chemical Properties External Links	Synonyms Produ	ust Composition Tax	Cest in Vitro Data	volure PubChem	Commen
				10- g			•		
				8-					
				7-			. ŝ.		
				6 6-			•		
		Το	wcast			1.1	1.1		
		10	xcust	3-		100			
		Car				· · ·	1 1 1 23	5. S. 8.	
		SCr	eening	0e0001 · · · · · · e0001 · · · ·	oði · · · ·	<u>a.</u> ī			ACRO MA
		-	. .						
		L	Jata	Download as: CSV Extel Sh	ow inactive				
				Assay Name	Hit Call 1	Top S	caled Top AC50	log AC50	Intended
				TOX21_AR_BLA_Antagonist_ratio	ACTIVE	106 2	26 29.5	1.47	nuclear re
				TOX21_AR_LUC_MDAKB2_Antag	onist ACTIVE	99.3 2	04 72.4	1.86	nuclear re
				TOY21 FRs RI & Annuist sh2	ACTIVE	121 2	10 0.747	JN 130	harkminin
Chemical Properties External Links	Synonyms Product Co	ToxCast in Vitro Data Ex Frequent Uses and Functions	φosure PubChem Comments						
building_mate	rial flooring	5							
building_constru	ction boats	6							
building_material flooring, building_ma	aterial walls	6							
manufacturi	ing plastics	6							
surface_treatm	nent metals	7			Eun	ction	nalll		ind
manufactu	ring metals		11		IUIN			JEU	IIU
manufacturing	, moonines				C				-
manufacturing	paint		13					IDa	ια
manufacturing	paint adhesive		13		Con	προε			
manufacturing	paint adhesive		13		Con	npos		1	
manufacturing	paint adhesive		13		Con (fron	n CPC	Cat)	
manufacturing	paintadhesive	Product Composition	13		Con ()	fron	n CPC	Cat)	
manufacturing Nownload as: CSV Excel	paint adhesive Percent	Product Composition	13 14 Manufacturer		Con (fron	n CPC	Cat)	
Download as: CSV Excel Product BISPHENOL-A (BPA)	paint adhesive Percent 100%	Product Composition Composition 4	13 14 Manufacturer GENERAL ELECTRIC COMP	ANY	Con (fron	n CPC	Cat)	
Download as: CSV Excel Product BISPHENOL-A (BPA) EPOXY PASTE PIGMENTS, 3402-	adhesive Percent 100% 3408 <100%	Product Composition Composition 4	13 14 Manufacturer GENERAL ELECTRIC COMP SYSTEM THREE RESINS	ANY	Con (fron	n CPC	Cat)	

The chemical page for Bisphenol A. Physical Properties Tab listing experimental and predicted properties.

						20182		۹	Options -
NC	CT Models:	Water S	olubility				Save Rep	ort	
Model R	esuits								
Predio Globa Local Confi	cted value: 3.68e-0 al applicability dom applicability doma dence level: 0.46 6	3 mol/L nain: Inside ain index: 0.	0 36 0						
Vi an				Weighted	KNN model		QMRF		
		5-fold CV (75%)	Trainin	g (75%)	Test (25%)			
		Q2	RMSE 0.81	R2	RMSE	R2	RMSE		
)	Ĵ		Ç	\mathfrak{O}	он				
halene	HO O	thalen-1-yi)p	2,3-Di	hydroxynapl	hthalene				
03 03	Measured: Predicted:	2.57e-02 1.38e-02	Me Pr	rasured: 1.83 edicted: 3.03	e-03 e-02				
ontact	town ACToR		Insuring DSI	STox	Privacy	Accessibility	Help		

Detailed Model Reports:

- Predicted data including logP, water solubility and **Bioconcentration Factor**
- Algorithms developed using data from PHYSPROP dataset
- Data curated using combined manual process & automated workflow¹
- Model Reports show global and local applicability domain and nearest neighbors in the training set
- "QSAR Modeling Report Format" (QMRF) reports

Non-Targeted Screening Support Through Mass and Formula-based Searching

• Support identification of chemicals in different environmental media (e.g. dust, water) utilizing mass spectrometry based non-targeted screening • Both mass- and formula-based searches are supported, single and batch • Utilize rank ordering of hit lists by the number of data sources, functional use and product use categorization

• Outperforms² community-standard database, ChemSpider, for compound identification. Demonstrated using peer-reviewed publication data

		A	Advanced Sea	arch	
Mass Sear	ch				
4	Min/Max				
Mess	amu	±	Error	amu	Search C
senerate N	noiecular Formula(e)			
Mass	amu	±	Error	amu	Search C
Include haloge	ms				
Onlines -					

Disruptive Innovation in Chemical Evaluation



Impacts

- Curated chemistry data: High quality curated chemistry datasets as foundation data for NCCT applications and available to other agency resources for reuse
- Flexible Platform Design: "Build-once, Use Many" approach provides architecture for other apps. RapidTox application already in development for internal use
- Integration Platform: Additional data streams can be added into the design to provide unifying integration platform for NCCT data and other agency resources

Coming Soon

- Batch Searching: Batch look-up and download of chemical names, CAS Numbers and DB mappings
- Literature Handling: Integrate to PubMed (with "Abstract Sifting", Google Scholar, Books and Patents
- **Predicted Functional Use**: CPDat database (replacing CPCat). Includes Predicted Functional Use
- iPhone/iPad App: Open Data allows reuse of the data

Future Work

- Support for ambiguous substances Data model will be expanded to include such ambiguous chemical substances as well as relationships within complex chemical mixtures
- Delivery of Public Web Services Application programming interface and web services to allow third parties to consume and integrate these data
- Access to real-time predictions Real-time prediction allowing scientists to submit one or more chemical structures and perform the predictions in real time

References

- 1. An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modeling, K. Mansouri, C.M. Grulke, A.M. Richard, R.S. Judson and A.J. Williams, Accepted for publication to SAR and QSAR in Environmental Research.
- 2. Identifying "known unknowns" using the US EPA's CompTox Chemistry Dashboard, A.D. McEachran, J.R. Sobus and A.J. Williams, Submitted for publication to Analytical and Bioanalytical Chemistry.

CSS BoSC Meeting 2016

EPA, ublic

2004454 Al nuclear recentor bacigonad resultante de Inorphology de Longo Silicit Annova de Longo Silicit Annova de Longo electrostatione de Longo ded Target Family
r receptor
'receptor
und mesonrement

Search Q Search Q Search Q