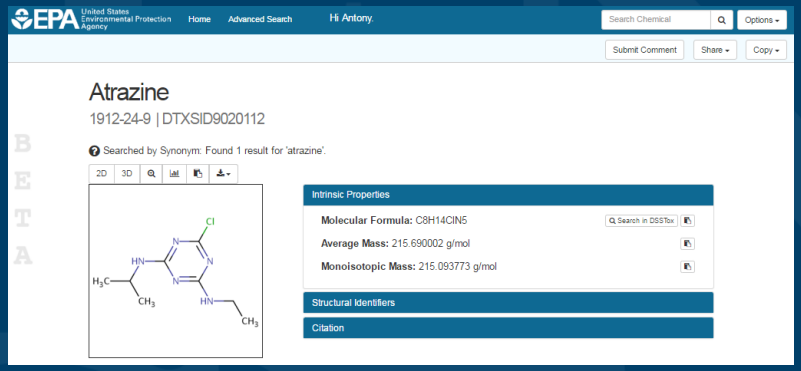
EPA United States Environmental Protection Agency

An Introduction to the iCSS Chemistry Dashboard



May 26, 2016

Antony Williams and Kamel Mansouri
National Center for Computational Toxicology

The iCSS Chemistry Dashboard

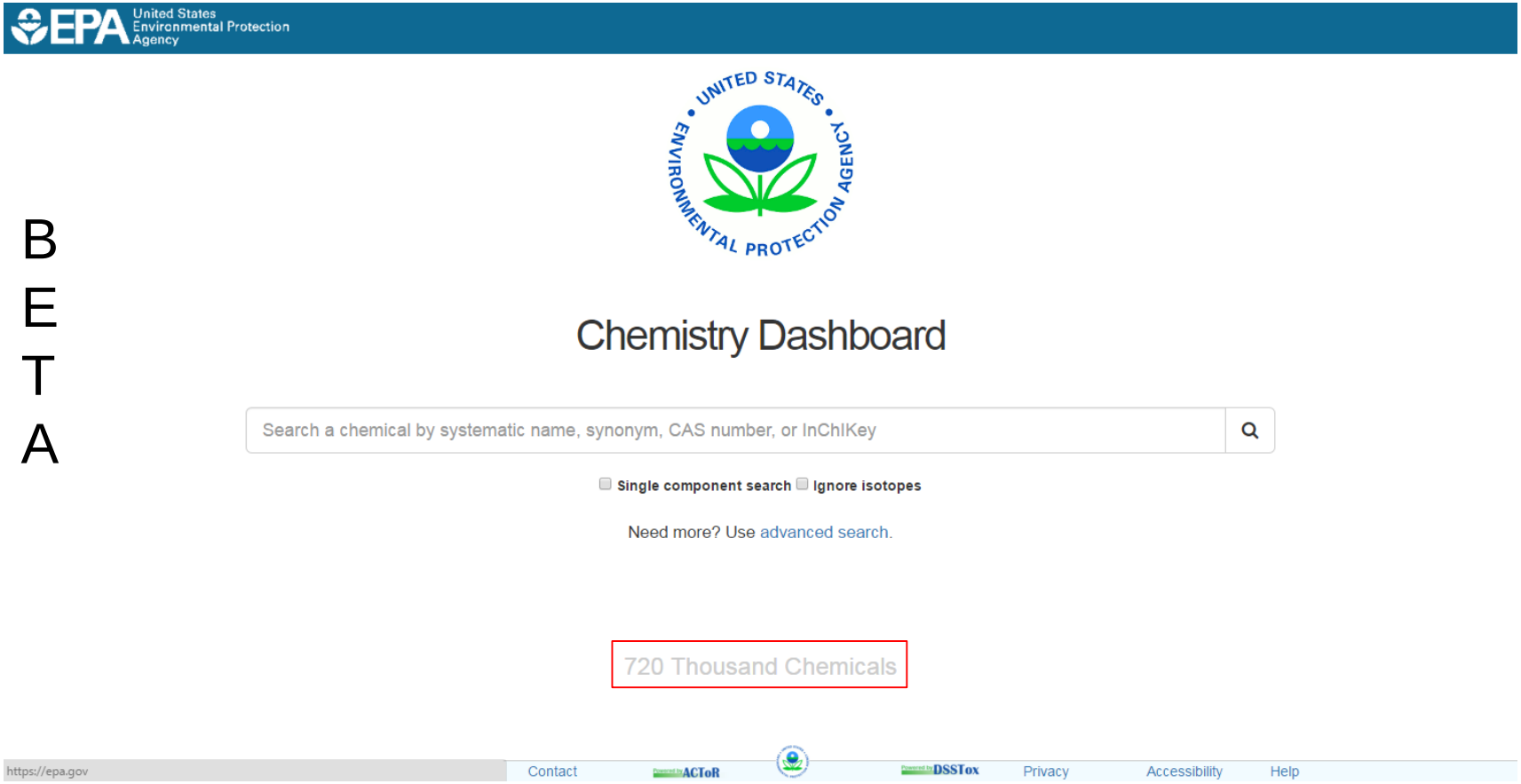
PRIMARY GOALS

- Deliver a web-based application serving up the chemistry related data used by our team
- Provide public access to the results of over a decade of curation work reviewing environmental chemistry data
- Provide access to the results of our QSAR modeling work
- Deliver a central hub to link together websites of interest
- All data to be available as Open Data for download/reuse

SECONDARY GOAL

- To develop a cheminformatics architecture to serve as a high quality chemical foundation for all NCCT tools and data

iCSS Chemistry Dashboard
https://comptox.epa.gov



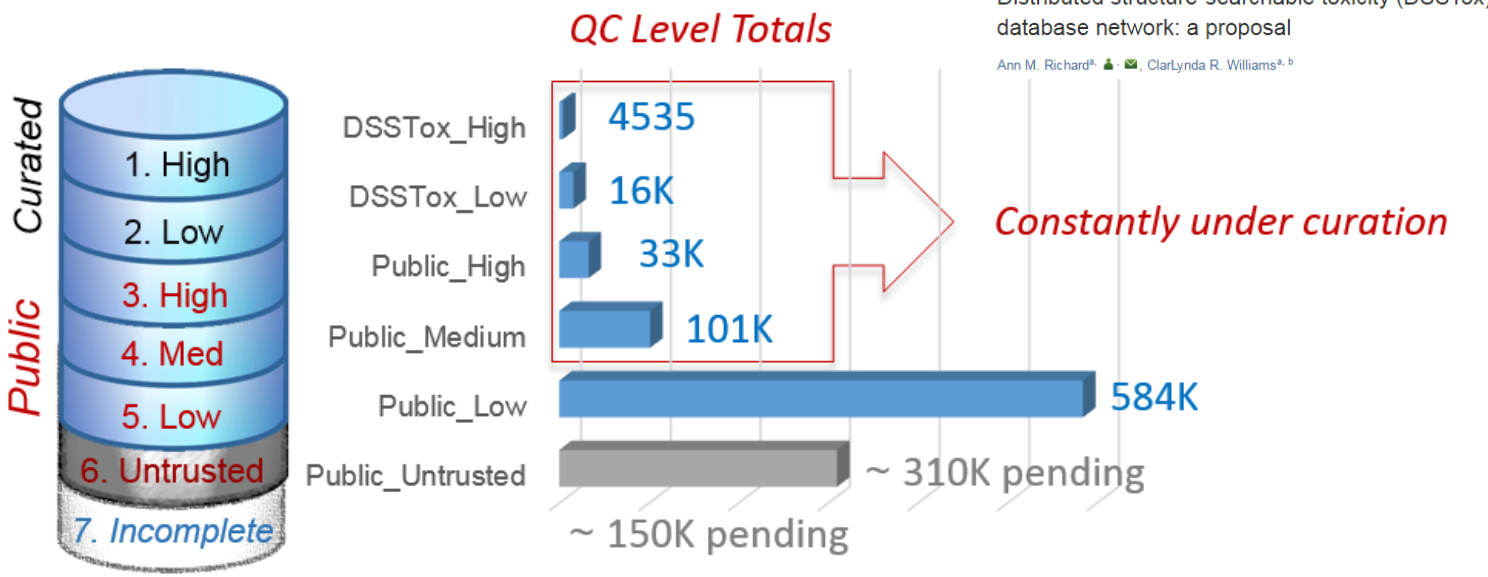
DSSTox Chemistry Content

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis
Volume 499, Issue 1, 29 January 2002, Pages 27–52

Mutation Research Frontiers

Distributed structure-searchable toxicity (DSSTox) public database network: a proposal

Ann M. Richard[a], ClarLynda R. Williams[a, b]



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

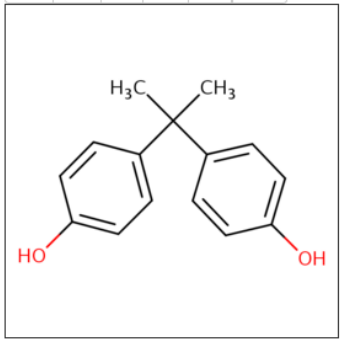
iCSS Chemistry Dashboard

Bisphenol A

80-05-7 | DTXSID7020182

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

2D 3D



Intrinsic Properties

Molecular Formula: C15H16O2 Search in DSSTox

Average Mass: 228.291 g/mol

Monoisotopic Mass: 228.11503 g/mol

Structural Identifiers

Citation

Chemical Properties | External Links | Synonyms | PubChem Biological Activities | PubChem Articles | PubChem Patents | Comments

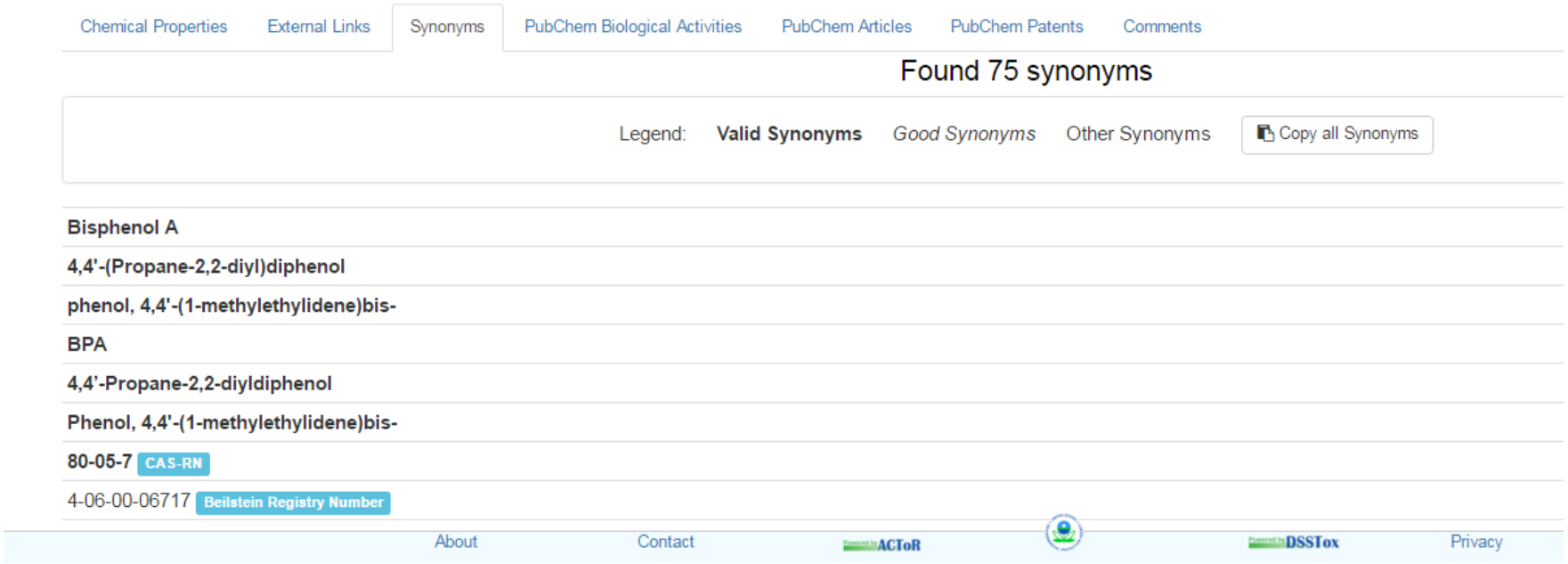
CSV Excel

Property	Average (Exp.)	Range (Exp.)	Average (Pred.)	Range (Pred.)
Solubility	0.001 (1)	0.0005257 to 0.0005257	0.38 (2)	0.003675 to 0.7565
Melting Point	154.929 (7)	153.0 to 158.0	144.033 (3)	131.8 to 158.0
Boiling Point	200.0 (1)	200.0 to 200.0	348.95 (2)	334.4 to 363.5
LogP	3.357 (3)	3.32 to 3.431	3.524 (3)	3.205 to 3.727
Atmospheric Hydroxylation Rate	N/A	N/A	0.0 (1)	4.237e-11 to 4.237e-11
LogBCF	1.64 (1)	1.64 to 1.64	1.376 (1)	1.376 to 1.376
Biodegradation Half-life	N/A	N/A	15.11 (1)	15.11 to 15.11
Henry's Law Constant	N/A	N/A	0.0 (1)	6.972e-07 to 6.972e-07

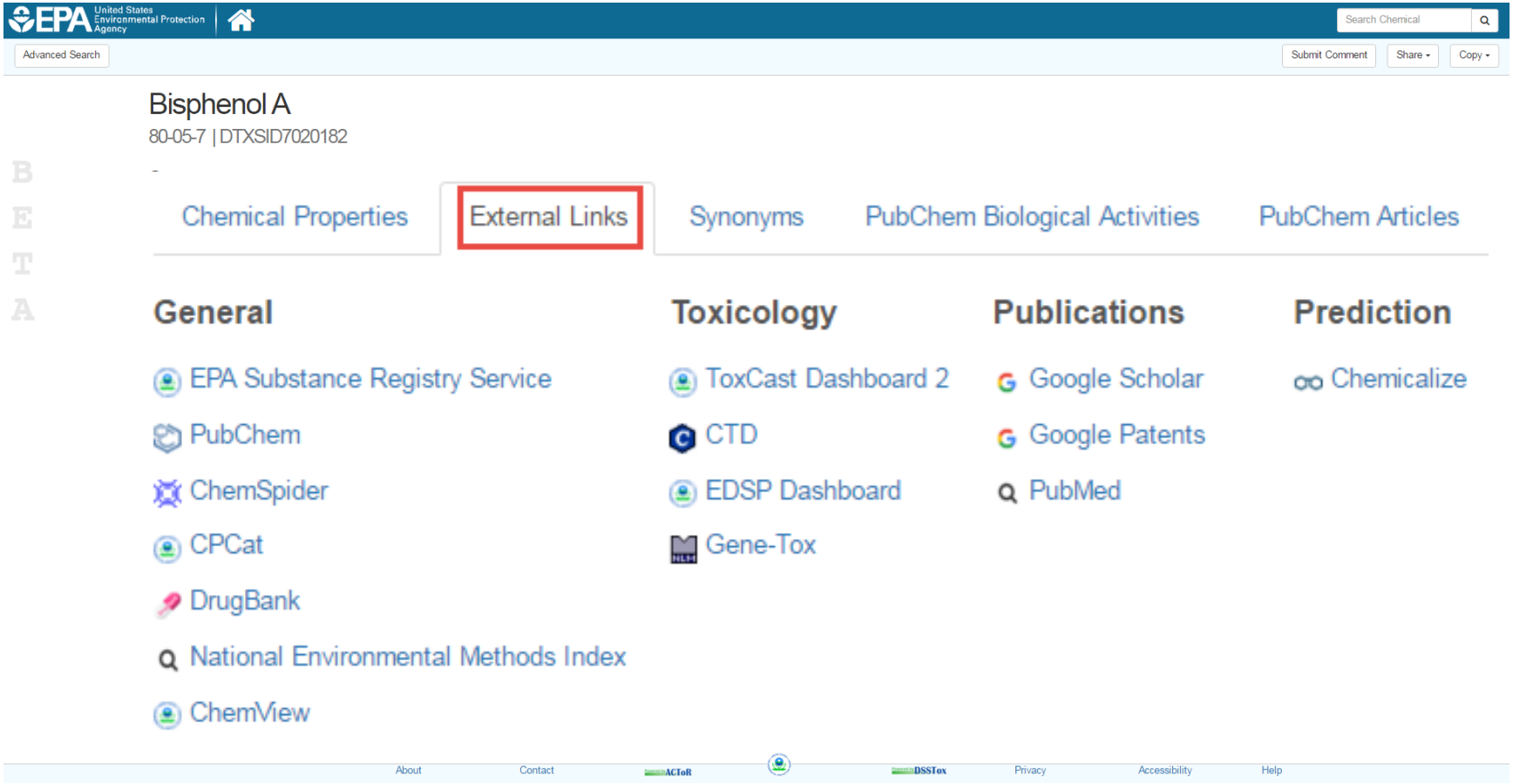
About Contact Powered by ACToR Powered by DSSTox Privacy Accessibility Help

>500,000 chemical names

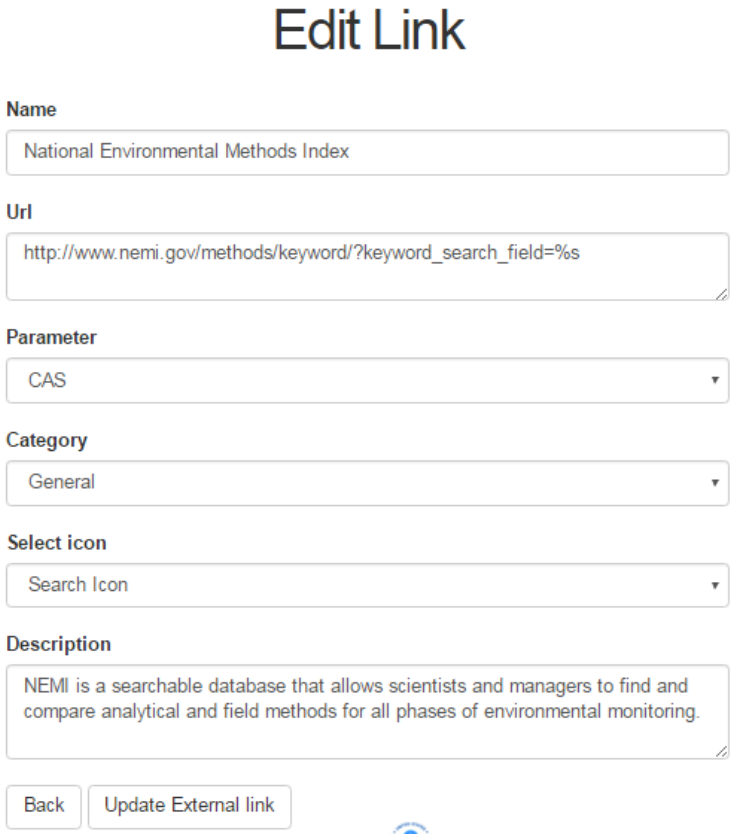
- Names, numeric identifiers and synonyms critical for:
  - –Searching database
  - –Integrating to external searches
  - –Supporting text-mining approaches

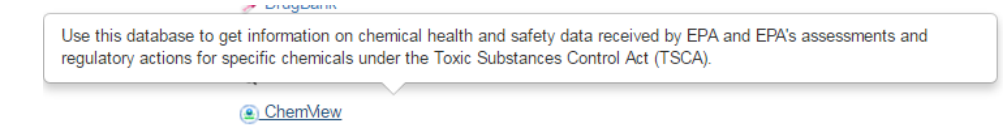


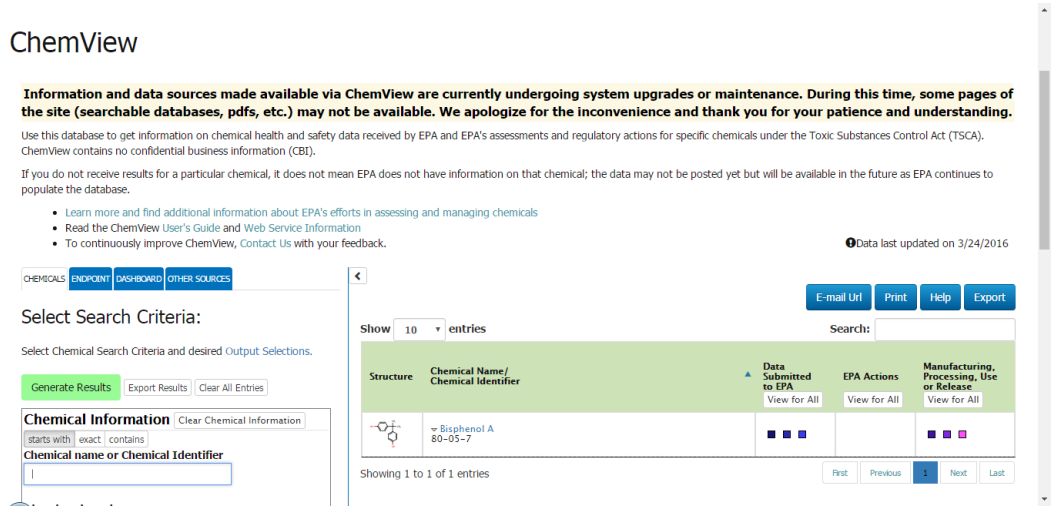
The iCSS Chemistry Dashboard



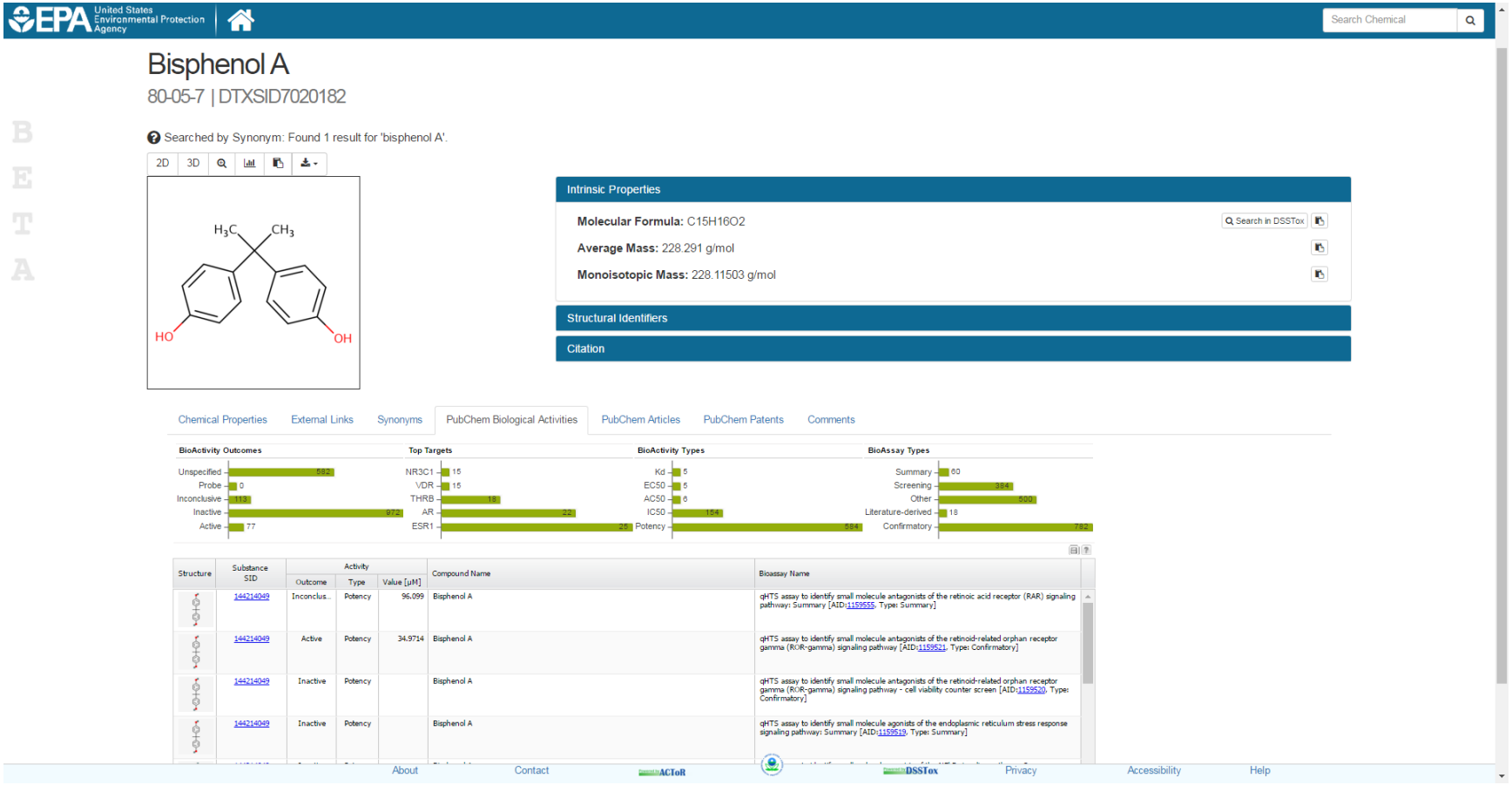
Make maintenance easy… Please suggest new ones!



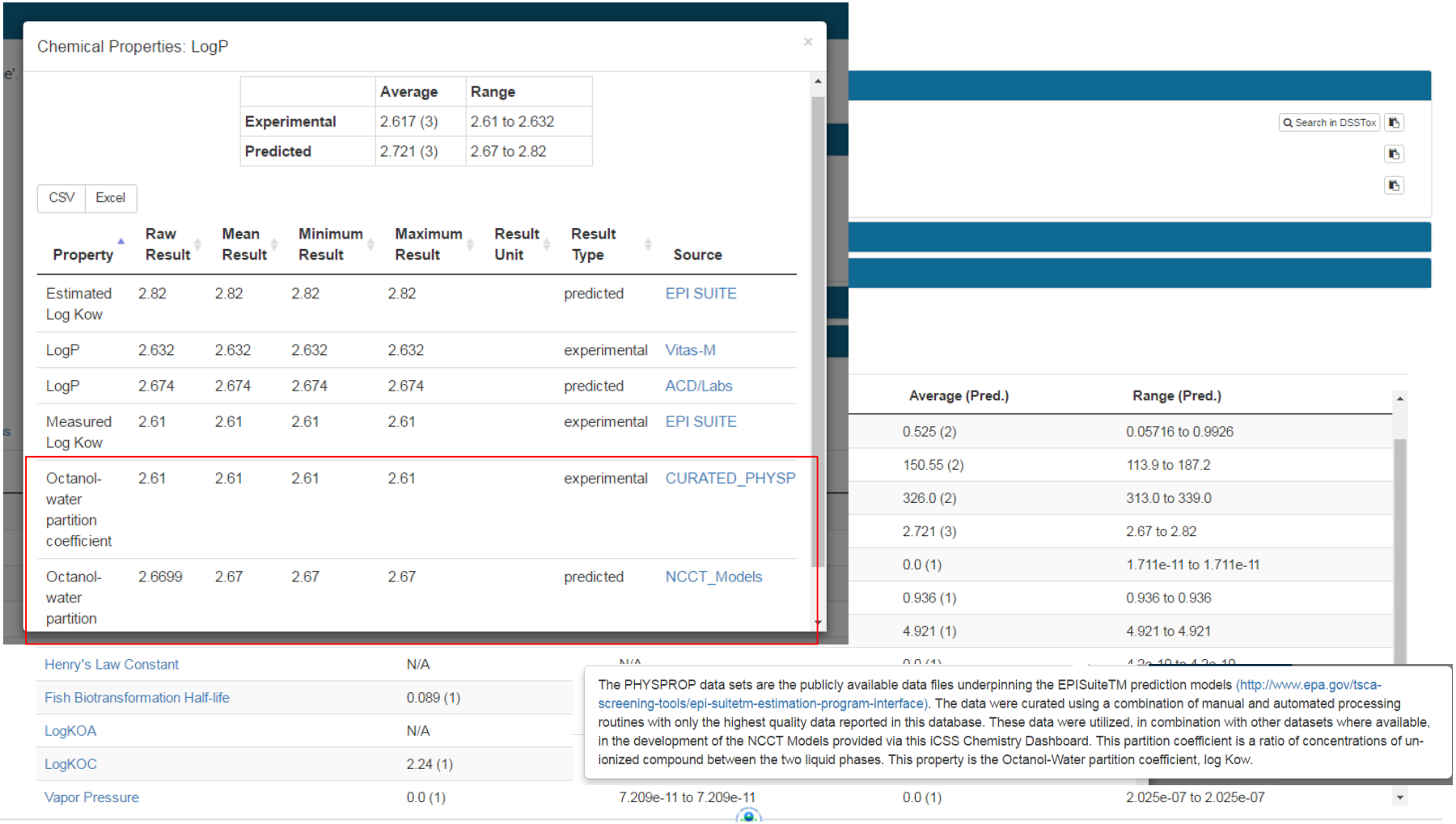




The iCSS Chemistry Dashboard



Expt'l and Predicted PhysChem Data



Data sharing and access

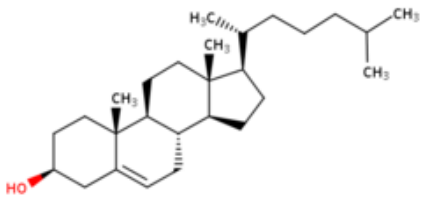
Cholesterol

57-88-5 | DTXSID3022401

Searched by Synonym: Found 1 result for 'cholesterol'.

2D | 3D | | | |

Mol File
Image



Chemical Properties: LogP

	Average	Range
Experimental	3.357 (3)	3.32 to 3.431
Predicted	3.524 (3)	3.205 to 3.727

CSV | Excel

Property	Raw Result	Mean Result	Minimum Result	Maximum Result	Result Unit	Result Type	Source
Estimated Log Kow	3.64	3.64	3.64	3.64		predicted	EPI SUITE
LogP	3.431	3.431	3.431	3.431		experimental	Vitas-M
LogP	3.727	3.727	3.727	3.727		predicted	ACD/Labs
Measured Log Kow	3.32	3.32	3.32	3.32		experimental	EPI SUITE
Octanol-water partition coefficient	3.32	3.32	3.32	3.32		experimental	CURATED_PHYSP

Found 75 synonyms

Legend: **Valid Synonyms** *Good Synonyms* Other Synonyms | Copy all Synonyms

Bisphenol A
4,4'-(Propane-2,2-diyl)diphenol
phenol, 4,4'-(1-methylethylidene)bis-
BPA
4,4'-Propane-2,2-diyldiphenol
Phenol, 4,4'-(1-methylethylidene)bis-
80-05-7 CAS-RN
4-06-00-06717 Beilstein Registry Number
UNII-MLT3645I99 FDA Registry Number
(4,4'-Dihydroxydiphenyl)dimethylmethane

New “NCCT Prediction Models”

- PhysChem properties to include in exposure modeling, augmenting with Toxcast data etc.
- Our approach to modeling:
  - Obtain high quality training sets
  - Apply machine learning approaches
  - Validate performance
  - Predict across full datasets

NCCT Model Development

- Dashboard access to cleaned data, the predicted data for all chemicals and ultimately real time models
- QSAR prediction models (kNN) produced for all properties
- >700k chemical structures pushed through NCCT_Models

EPA United States Environmental Protection Agency

Available Properties

- Solubility
- Melting Point
- Boiling Point
- LogP (KOWWIN: Octanol-water partition coefficient)
- Atmospheric Hydroxylation Rate
- LogBCF (Bioconcentration Factor)
- Biodegradation Half-life
- Henry's Law Constant
- Fish Biotransformation Half-life
- LogKOA (Octanol/Air Partition Coefficient)
- LogKOC (Soil Adsorption Coefficient)
- Vapor Pressure

Some statistics

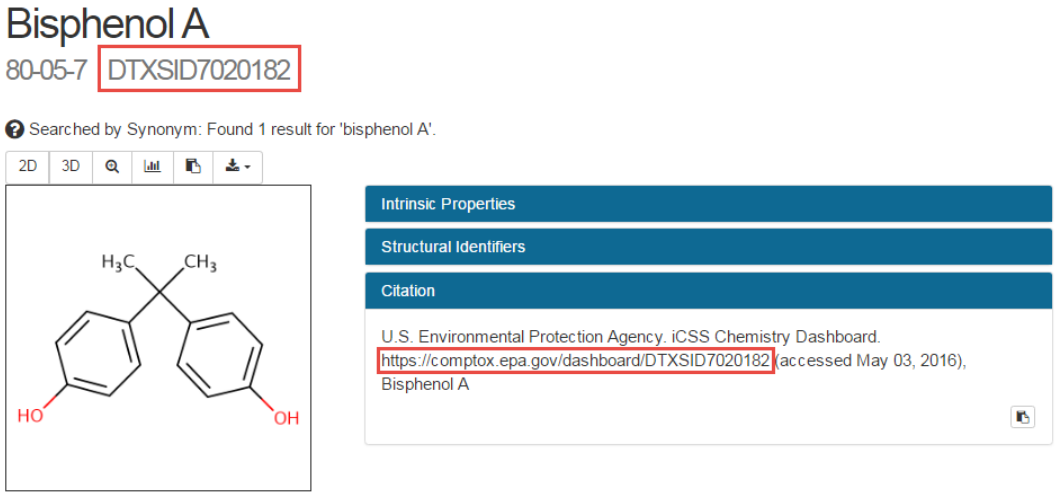
- 15,809 chemicals in the LogP dataset
- CAS Checksum: 12163 valid, 3646 invalid **(>23%)**
- Valence errors: 322 Molfile vs 3782 SMILES **(>24%)**
- SMILES vs. Molfiles (structure check)
  - –133 SMILES cannot be converted **(~1%)**
  - –1279 differ in stereochemistry **(~8%)**
  - –362 “Covalent Halogens”
  - –191 differ as “tautomers”
  - –436 are different compounds **(~3%)**
- **Original Training Set: >2400; Curated Training Set > 14000**

Adding MORE Data

- Collect “Open Data” for various endpoints from:
  - *PubChem*
  - *Open Data Sources*
- Data already harvested and awaiting processing
- New models to be produced from expanded data sets

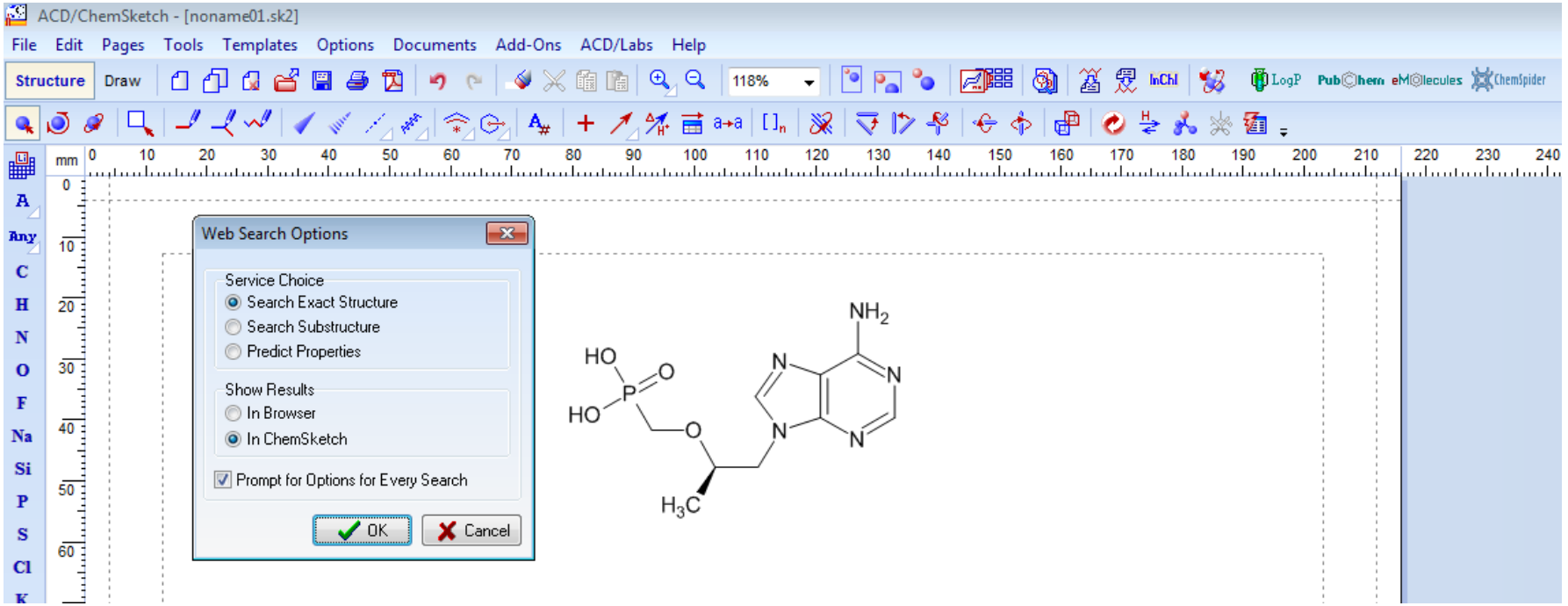
Connecting into the Dashboard

- Linkages into the Dashboard are simple: using the associated identifiers



- For integration we can supply files of structures/identifiers mapped to DTXSIDs. Contact us…
- PubChem, EBI's UNICHEM, ChemSpider and…

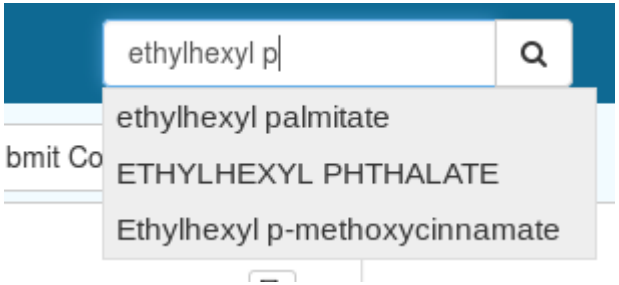
Desktop integration coming



Next Release in June..

- June release delivery update
  - –Faster searching
  - –Improved navigation
  - –New data types – integration to other NCCT apps
    - *In vitro* DB
    - Chemical and Product categories (CPCat)
    - Exposure Data
  - –Expanded list of External Links
  - –Calculation details for models

Type Ahead Search



New Table Displays

Chemical Properties | External Links | Synonyms | Consumer Usage | ToxCast in Vitro Data | Exposure | Analytical | PubChem

Summary

- Solubility
- Melting Point
- Boiling Point
- LogP
- Atmospheric Hydroxylation Rate
- LogBCF
- Biodegradation Half-life
- Henry's Law Constant
- Fish Biotransformation Half-life
- LogKOA
- LogKOC
- Vapor Pressure

CSV | Excel

Property	Average (Exp.)	Median (Exp.)	Range (Exp.)	Average (Pred.)	Median (Pred.)	Range (Pred.)	Result Unit
Solubility	0.026 (1)	0.0255	0.0255	14.8 (2)	14.8	0.146 to 29.4	mol/L
Melting Point	136 (13)	137	122 to 142	109 (2)	109	93.1 to 125	°C
Boiling Point	140 (2)	140	140	297 (2)	297	278 to 316	°C
LogP	1.19 (2)	1.19	1.19	1.33 (3)	1.41	1.13 to 1.44	-
Atmospheric Hydroxylation Rate	-	0	-	0 (1)	3.396e-11	3.396e-11	cm3/molecule-sec
LogBCF	-	0	-	0.619 (1)	0.619	0.619	-
Biodegradation Half-life	-	0	-	3.54 (1)	3.54	3.54	days
Henry's Law Constant	-	0	-	0 (1)	3.095e-05	3.095e-05	atm-m3/mole
Fish		0		0.114 (1)	0.114	0.114	days

New Table Displays

Chemical Properties | External Links | Synonyms | Consumer Usage | ToxCast in Vitro Data | Exposure | Analytical | PubChem

- Summary
- Solubility
- Melting Point
- Boiling Point
- **LogP**
- Atmospheric Hydroxylation Rate
- LogBCF
- Biodegradation Half-life
- Henry's Law Constant
- Fish Biotransformation Half-life
- LogKOA
- LogKOC
- Vapor Pressure

LogP			
	Average	**Median**	**Range**
Experimental	1.19 (2)	1.19	1.19
Predicted	1.33 (3)	1.41	1.13 to 1.44

CSV | Excel

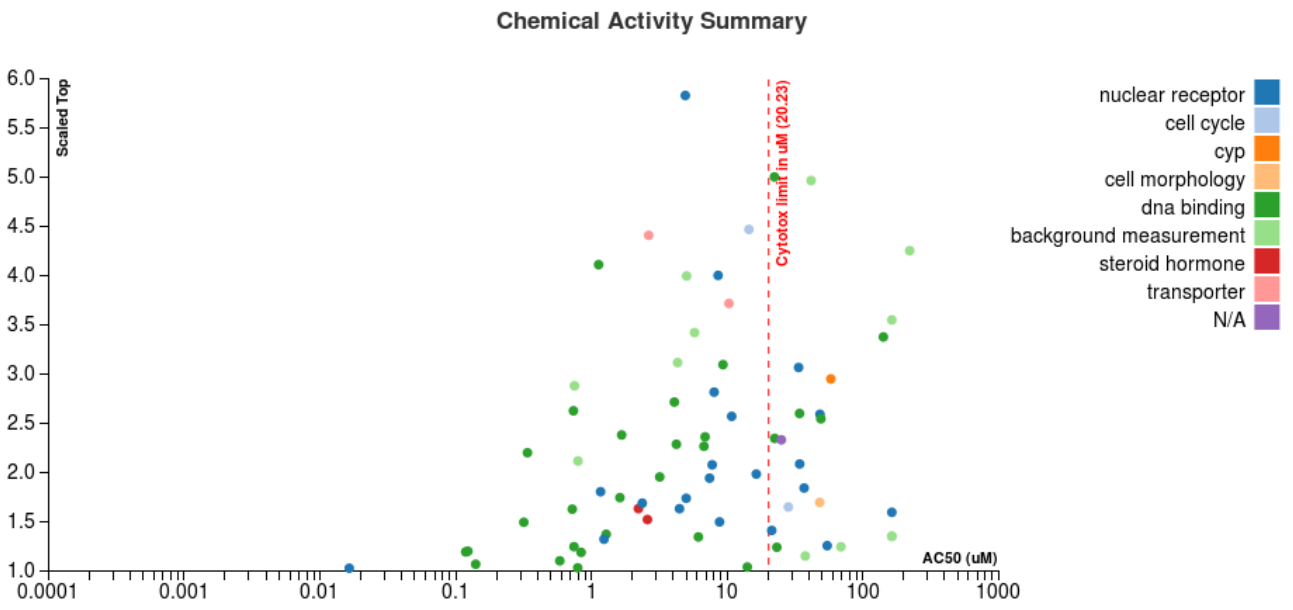
Experimental		
Source	**Result**	
EPI SUITE	1.19	-
CURATED PHYSPROP	1.19	-

Predicted		
Source	**Result**	**Calculation Details**
EPI SUITE	1.13	
ACD/Labs	1.44	
NCCT Models	1.41	

Calculation Details link to reports for T.E.S.T data and NCCT_Models

In vitro DB…

Chemical Properties | External Links | Synonyms | Consumer Usage | ToxCast in Vitro Data | Exposure | Analytical | PubChem | Comments



Assay Name	Hit Call	Top	Scaled Top	AC50	log AC50	Intended Target Family
ATG_TRANS	ACTIVE	1.680	1.419	21.317	1.329	nuclear receptor
TOX21_AR_BLA_Antagonist	ACTIVE	98.040	2.086	7.763	0.890	nuclear receptor

Consumer Usage (CPCat)

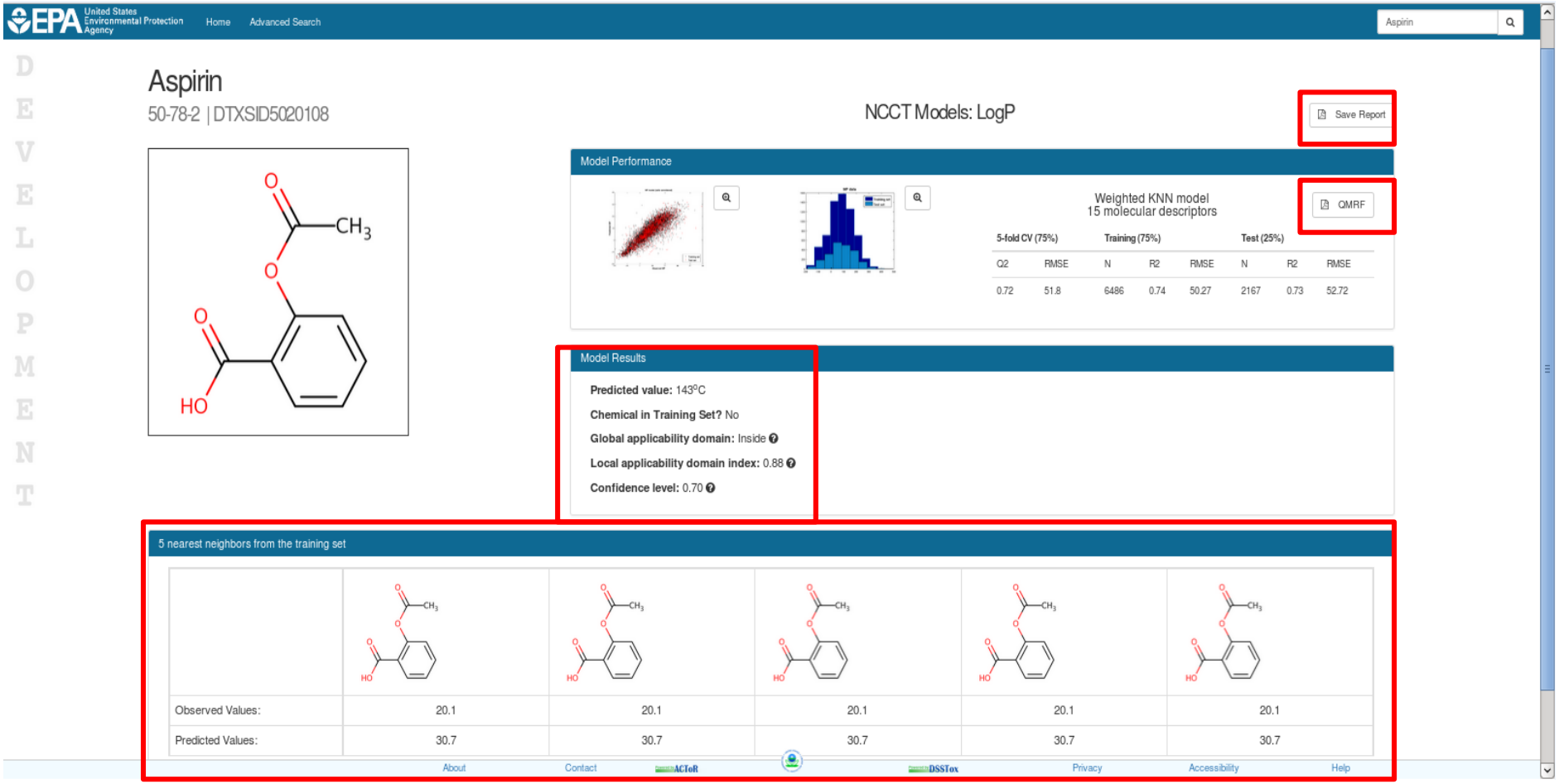
Chemical Properties | External Links | Synonyms | Consumer Usage | ToxCast in Vitro Data | Analytical Comments | Toxicity Values | PubChem

Frequent Uses and Functions

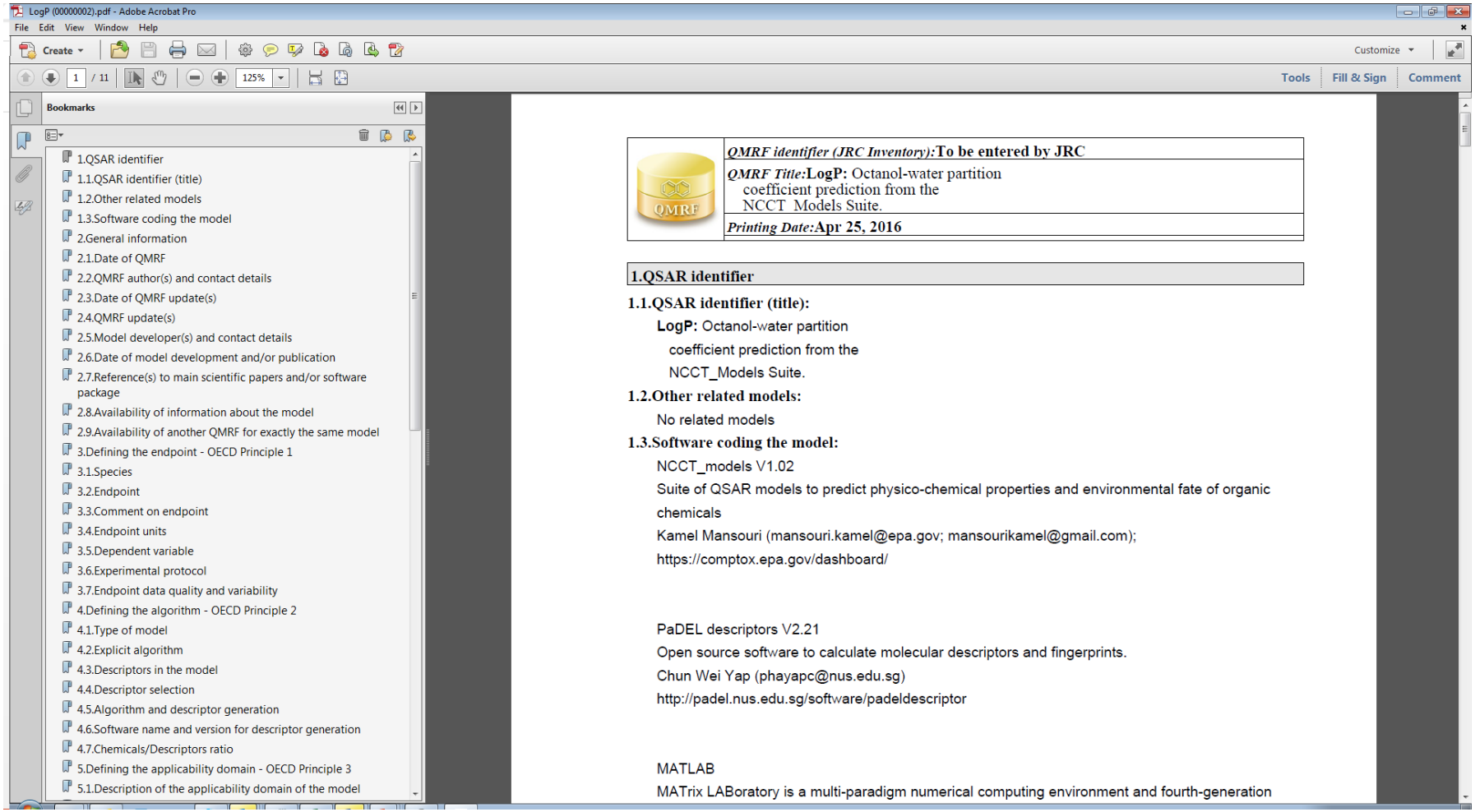
Product Composition

Product	Percent Composition	Manufacturer
CALCOZINE AURAMINE 00		AMERICAN CYANAMID CO
B604 AURAMINE O (C I 41000)		J T BAKER INC
AURAMINE O, B604	90-100%	J T BAKER INC
TB AURAMINE RHODAMINE T 3317	1.2%	DIFCO LABORATORIES,INC.
RET-SEARCH DYE/REC-800		SYSMEX

Calculation Details



QSAR Modeling Reporting Format



Toxicity Estimation Software Tool

US Environmental Protection Agency

Español | 中文: 繁體版 | 中文: 简体版 | Tiếng Việt | 한국어

Learn the Issues | Science & Technology | Laws & Regulations | About EPA

Search EPA.gov

Related Topics: Safer Chemicals Research

Contact Us Share

Toxicity Estimation Software Tool (TEST)

On this page:

- QSAR Methodologies
- New! What's New in Version 4.2?
- Prior Version History
- System Requirements
- Installation Instructions

The software includes models for the following endpoints:

- 96-hour fathead minnow 50 percent lethal concentration (LC50)
- 48-hour daphnia magna 50 percent lethal concentration (LC50)
- Tetrahymena pyriformis 50 percent growth inhibition concentration (IGC50) Exit
- Oral rat 50 percent lethal dose (LD50) Exit
- Bioconcentration Factor (BCF) The bioconcentration factor data set was compiled by researchers at the Mario Negri Istituto Di Ricerche Farmacologiche Exit
- Developmental Toxicity (DevTox) Exit
- Ames Mutagenicity (Mutagenicity) Exit

Subset of T.E.S.T Reports

View PDF

Predicted Water solubility at 25°C for 75-05-8 from Consensus method

Prediction results

Endpoint	Experimental value (CAS= 75-05-8) Source: EPI Suite v 4.00	Predicted value[a]
Water solubility at 25°C -Log10(mol/L)	-1.39	0.06
Water solubility at 25°C mg/L	1000965.12	35541.38

[a]Note: the test chemical was present in the training set. The prediction *does not* represent an external prediction.

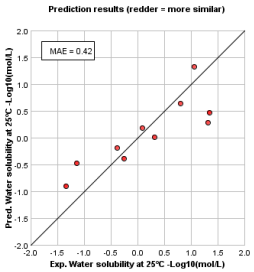
Individual Predictions

Method	Predicted value -Log10(mol/L)
Hierarchical clustering	N/A
Group contribution	0.34
Nearest neighbor	-0.21

Test chemical

Predictions for the test chemical and for the most similar chemicals in the external test set

If *similar* test set chemicals were predicted well relative to the entire test set, one has greater confidence in the predicted value.

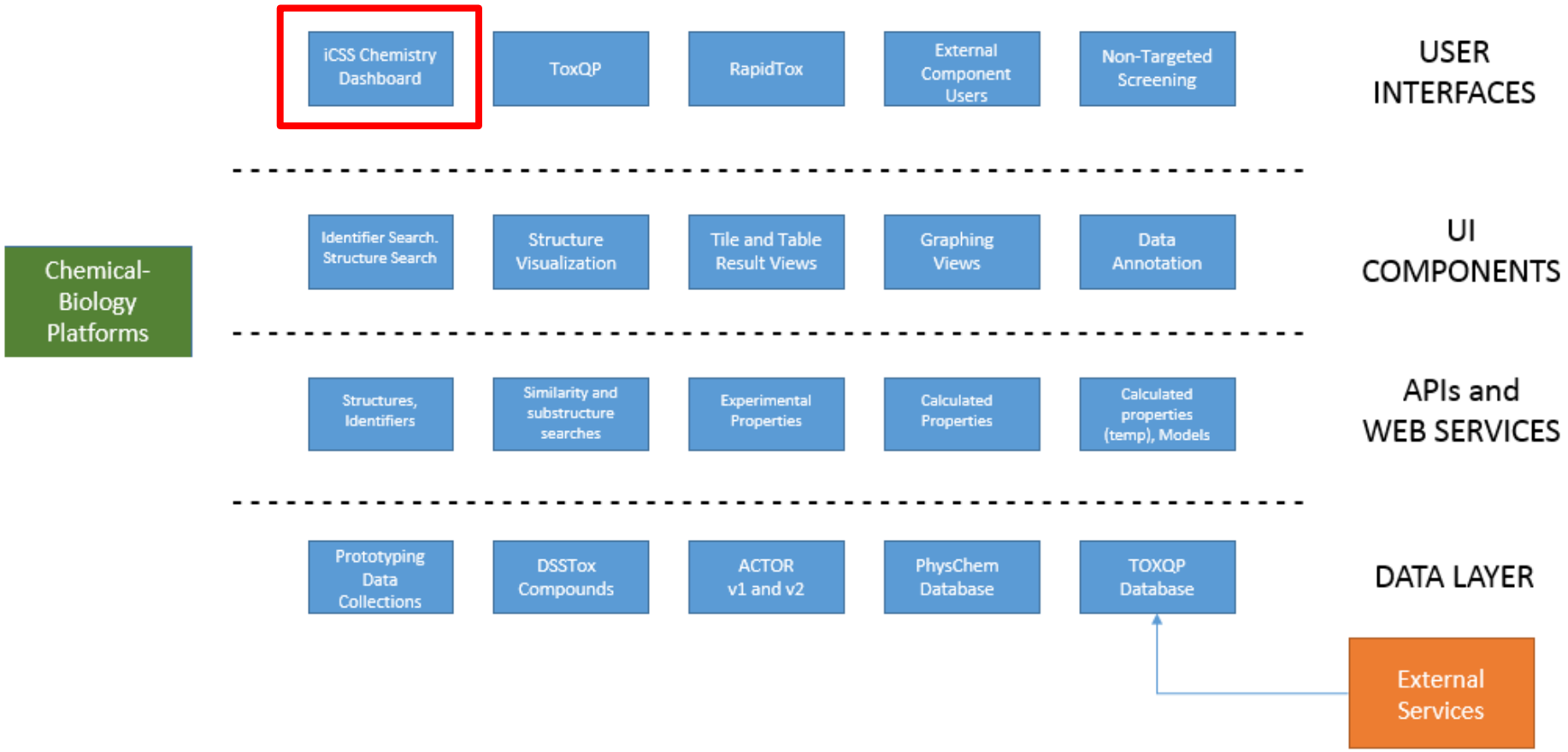


Chemicals	MAE*
Entire set	0.58
Similarity coefficient ≥ 0.5	0.42

*Mean absolute error in -Log10(mol/L)

CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/L)	Predicted value -Log10(mol/L)
75-05-8 (test chemical)			-1.39	0.06
64-17-5		0.92	-1.34	-0.91
353-36-6		0.92	1.35	0.48

Chemistry Software Architecture



Summary

- iCSS Chemistry Dashboard has been in public *beta* for 2 months. Version 1 release scheduled for June 2016
- Validation of developing cheminformatics architecture

- Future Work:
  - Provide programmatic access to ORD data and algorithms
  - Integrate more ORD resources – e.g. ECOTOX, T.E.S.T
  - Integrate additional data resources
  - Integrate with NCCT applications
  - Support for chemical substances without structures
  - Continue to enhance our **"*build once, use many*"** architecture

EPA United States Environmental Protection Agency

Acknowledgments

Contact Information

Dashboard: Antony Williams

williams.antony@epa.gov

919-541-1033

NCCT_Models: Kamel Mansouri

Mansouri.kamel@epa.gov

919-541-0545