EPA
United States
Environmental Protection
Agency

Prediction of toxicity and comparison of alternatives using WebTEST (Webservices Toxicity Estimation Software Tool)

Todd Martin, EPA/NRMRL
Antony Williams, EPA/NCCT
Valery Tkachenko, Science Data Experts

Office of Research and Development
National Risk Management Research Laboratory / Sustainable Technology Division / Clean Processes Branch

June 18, 2018

EPA
United States
Environmental Protection
Agency

Disclaimer

- The findings and conclusions in this presentation have not been formally disseminated by the U.S. EPA and should not be construed to represent any agency determination or policy.

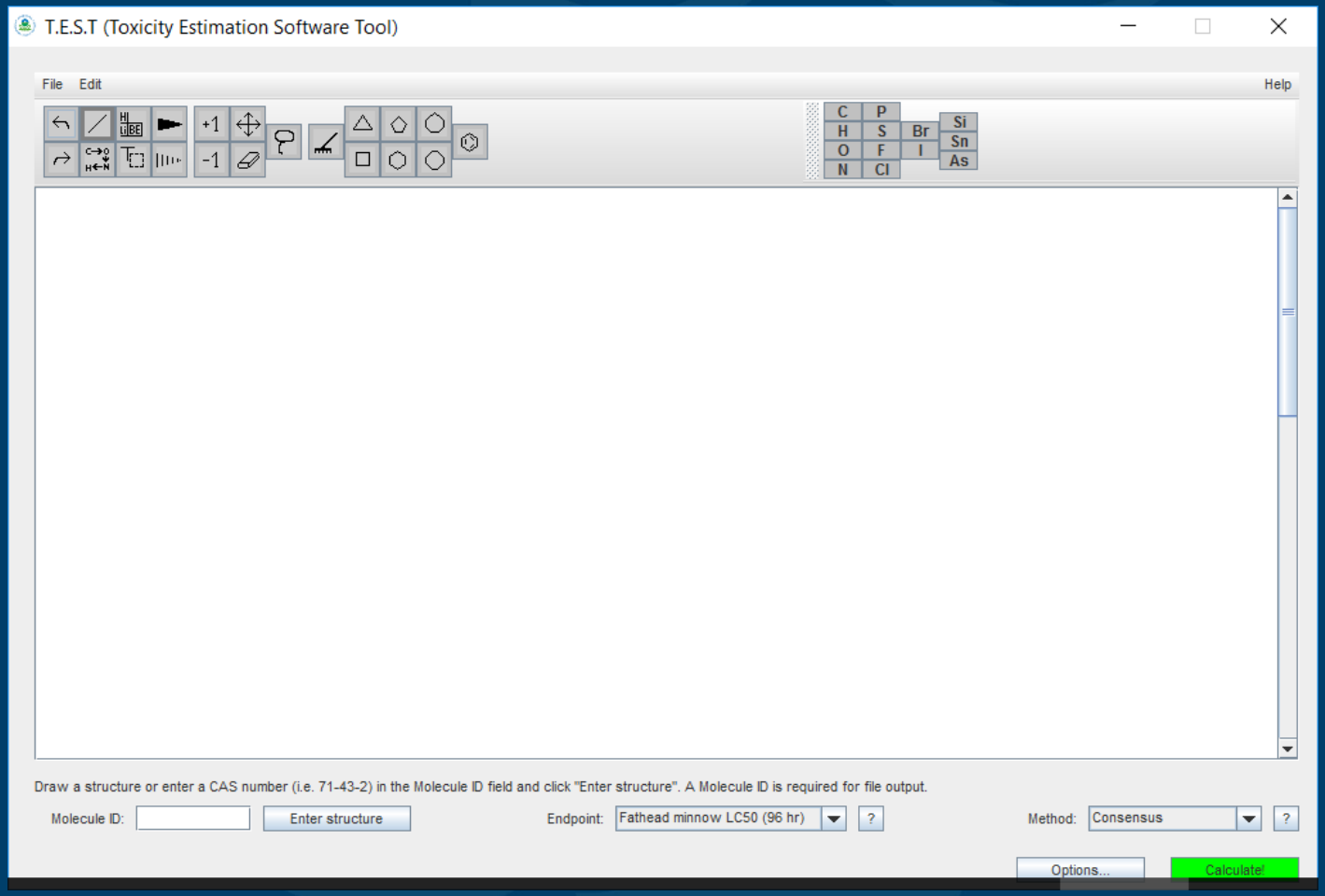
Abstract

- A Java-based web service is being developed within the US EPA's Chemistry Dashboard to provide real time estimates of toxicity values and physical properties. WebTEST can generate toxicity predictions directly from a simple URL which includes the endpoint, QSAR method, and SMILES string. An API has been developed to allow prediction of toxicity from GET and POST commands. A new web interface has been created to allow users to make predictions for single chemicals (by drawing chemicals in a Ketcher chemical structure editor or by searching for chemicals in the US EPA's Chemistry Dashboard). Previously calculated results for over 700 thousand chemicals are stored in a database to improve response time.

- The web service will also generate alternative assessment (AA) hazard profiles to compare chemical alternatives. Alternatives assessment frameworks evaluate chemical alternatives in terms of human health effects, ecotoxicity, and fate. Example categories include acute mammalian toxicity, acute aquatic toxicity, and bioaccumulation, respectively. Online data sources such as Japan's National Institute of Technology (NITE) can be utilized to obtain GHS (Global Harmonization System) scores for comparing alternatives. Data gaps can be filled using the toxicity models within WebTEST (e.g. the oral rat LD50 model)

History

- In 2008 T.E.S.T. was released as a Java application:
  - https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test



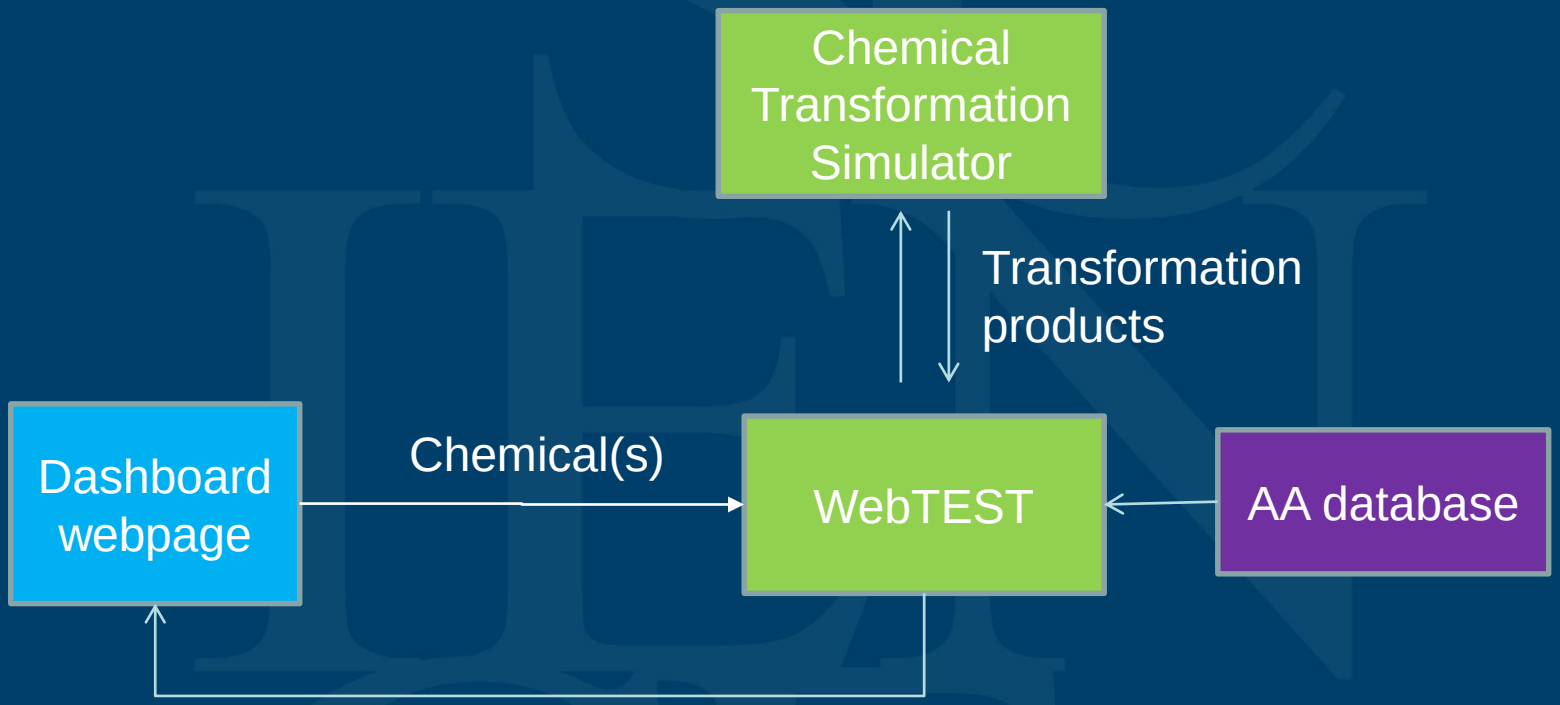
History, cont.

- In 2017, work began to convert T.E.S.T. to a web-services based application (WebTEST)

- WebTEST was used to make predictions for 743,000 chemicals which were then made available on the EPA's Chemistry Dashboard:
  - https://comptox.epa.gov/dashboard

- Real time predictions can be accessed via the predictions tab on the Chemistry Dashboard:
  - https://comptox.epa.gov/dashboard/predictions/index

Outline of WebTEST



.json / .html

1. Predict toxicity
2. Alternatives assessment
3. Sustainable synthesis

"GET" API Call

URL/endpointAbbreviation?smiles=desiredSmiles&method=methodAbbreviation

where URL = https://comptox.epa.gov/dashboard/web-test/

Endpoint	Abbreviation
Fathead minnow LC50 (96 hr.)	LC50
Daphnia magna LC50 (48 hr.)	LC50DM
T. pyriformis IGC50 (48 hr.)	IGC50
Oral rat LD50	LD50
Bioaccumulation factor	BCF
Developmental Toxicity	DevTox
Mutagenicity	Mutagenicity
Normal boiling point	BP
Vapor pressure at 25° C	VP
Melting point	MP
Flash point	Density
Density	FP
Surface tension at 25° C	ST
Thermal conductivity at 25° C	TC
Viscosity at 25° C	Viscosity
Water solubility at 25° C	WS

Method	Abbreviation
Hierarchical clustering	hc
Single model	sm
Nearest neighbor	nn
Group contribution	gc
Consensus	consensus (default)

Example “GET” Call

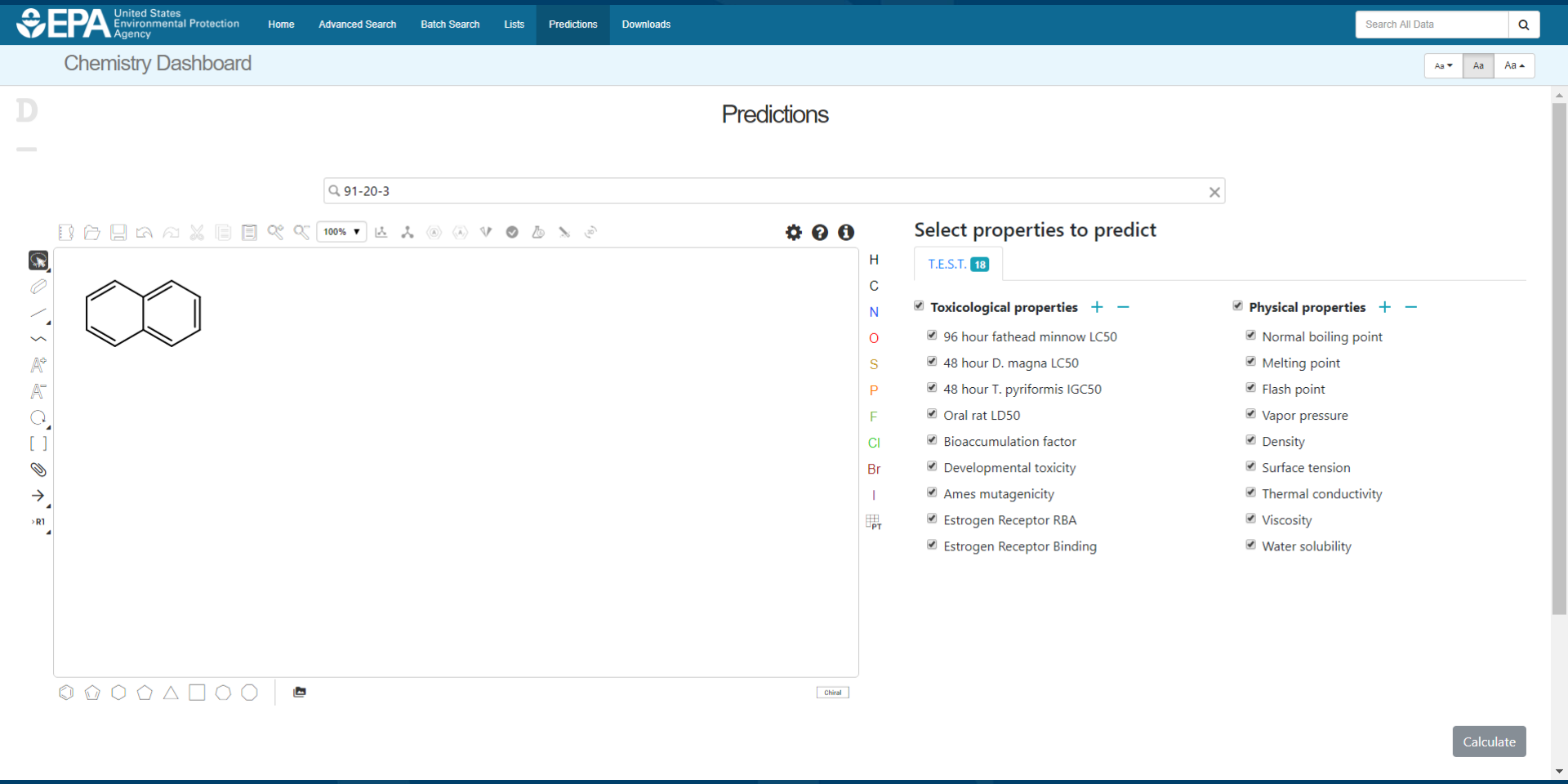
https://comptox.epa.gov/dashboard/web-test/WS?smiles=CCO&method=hc

JSON Raw Data Headers

Save Copy

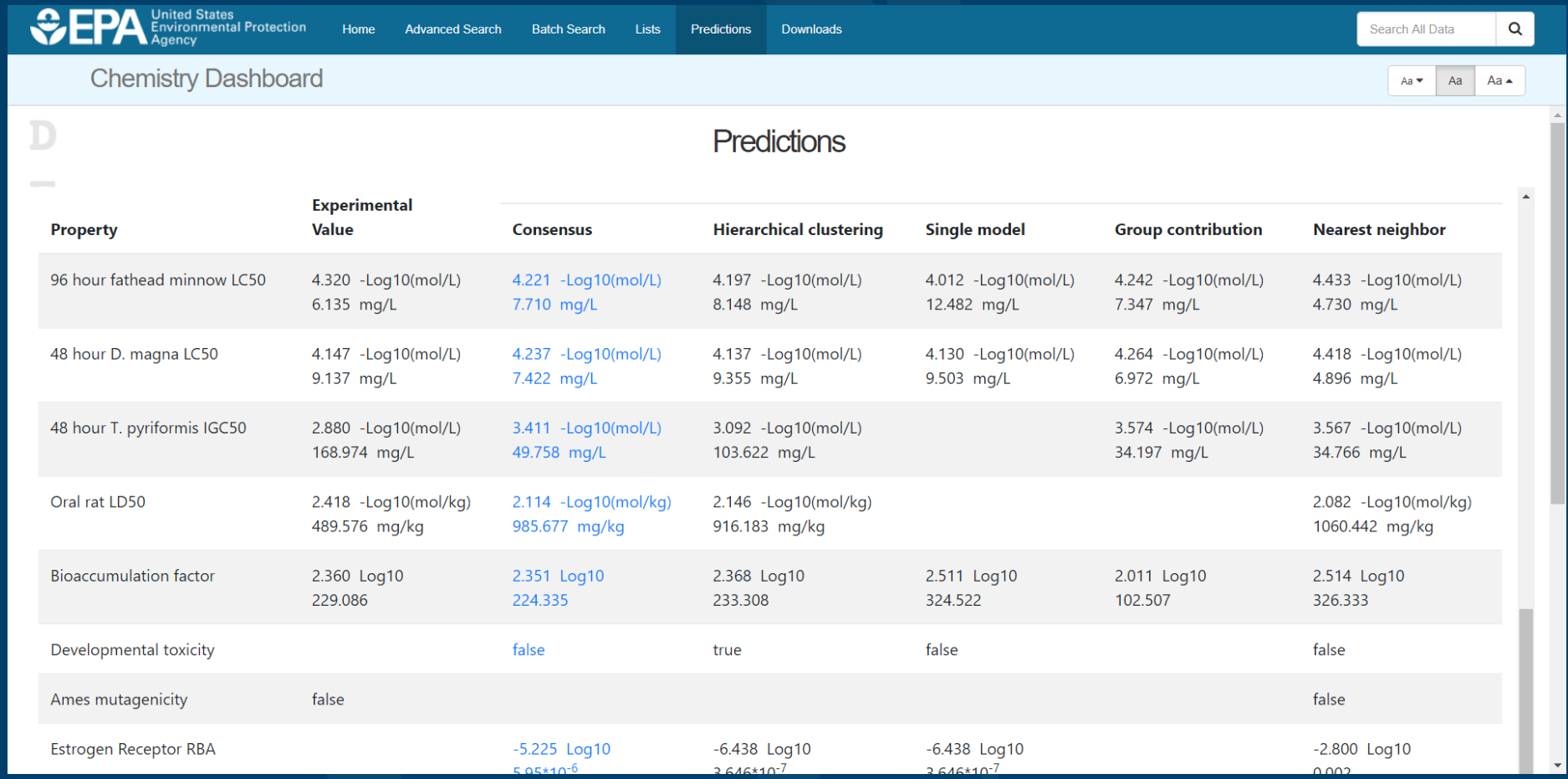

```
uuid:                "55547f4f-f966-48e8-b831-a0d217998064"
predictionTime:      1520539090089
software:            "T.E.S.T (Toxicity Estimation Software Tool)"
softwareVersion:     "5.01"
condition:           "25°C"
endpoint:            "Water solubility at 25°C"
method:              "Hierarchical clustering"
predictions:
  0:
    id:              "C_1520539090089"
    smiles:          "OCC"
    expValMolarLog:  "-1.337"
    expValMass:      "1001180.703"
    predValMolarLog: "-1.338"
    predValMass:     "1002625.241"
    molarLogUnits:   "-Log10(mol/L)"
    massUnits:       "mg/L"
```

WebTEST predictions on the Chemistry Dashboard



https://comptox.epa.gov/dashboard/predictions/index

WebTEST output



EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search | Lists | Predictions | Downloads

Search All Data

Chemistry Dashboard

Predictions

Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50	4.320 -Log10(mol/L) 6.135 mg/L	4.221 -Log10(mol/L) 7.710 mg/L	4.197 -Log10(mol/L) 8.148 mg/L	4.012 -Log10(mol/L) 12.482 mg/L	4.242 -Log10(mol/L) 7.347 mg/L	4.433 -Log10(mol/L) 4.730 mg/L
48 hour D. magna LC50	4.147 -Log10(mol/L) 9.137 mg/L	4.237 -Log10(mol/L) 7.422 mg/L	4.137 -Log10(mol/L) 9.355 mg/L	4.130 -Log10(mol/L) 9.503 mg/L	4.264 -Log10(mol/L) 6.972 mg/L	4.418 -Log10(mol/L) 4.896 mg/L
48 hour T. pyriformis IGC50	2.880 -Log10(mol/L) 168.974 mg/L	3.411 -Log10(mol/L) 49.758 mg/L	3.092 -Log10(mol/L) 103.622 mg/L		3.574 -Log10(mol/L) 34.197 mg/L	3.567 -Log10(mol/L) 34.766 mg/L
Oral rat LD50	2.418 -Log10(mol/kg) 489.576 mg/kg	2.114 -Log10(mol/kg) 985.677 mg/kg	2.146 -Log10(mol/kg) 916.183 mg/kg			2.082 -Log10(mol/kg) 1060.442 mg/kg
Bioaccumulation factor	2.360 Log10 229.086	2.351 Log10 224.335	2.368 Log10 233.308	2.511 Log10 324.522	2.011 Log10 102.507	2.514 Log10 326.333
Developmental toxicity		false	true	false		false
Ames mutagenicity	false					false
Estrogen Receptor RBA		-5.225 Log10 $5.95*10^{-6}$	-6.438 Log10 $3.646*10^{-7}$	-6.438 Log10 $3.646*10^{-7}$		-2.800 Log10 0.002

Prediction report

Predicted Fathead minnow LC50 (96 hr) for 91-20-3 from Consensus method

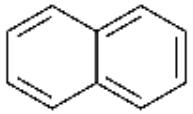
Prediction results

Endpoint	Experimental value (CAS= 91-20-3) Source: ECOTOX	Predicted value[a]
Fathead minnow LC_{50} (96 hr) -Log10(mol/L)	4.32	4.22
Fathead minnow LC_{50} (96 hr) mg/L	6.14	7.71

[a]Note: the test chemical was present in the external test set.

Individual Predictions

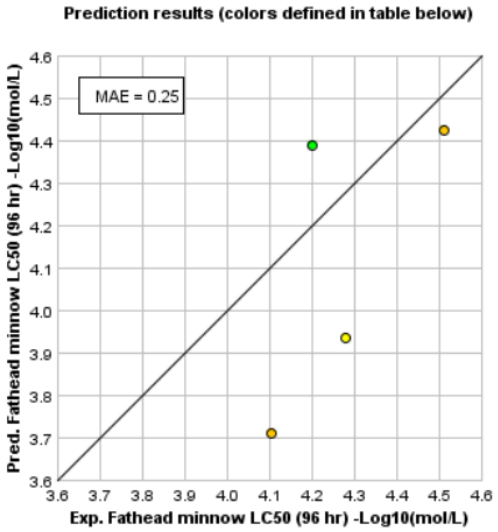
Method	Predicted value -Log10(mol/L)
Hierarchical clustering	4.20
Single model	4.01
Group contribution	4.24
Nearest neighbor	4.43



Prediction report cont.

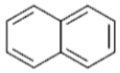
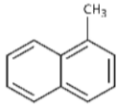
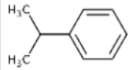
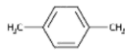
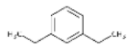
Predictions for the test chemical and for the most similar chemicals in the external test set

If the predicted value matches the experimental values for similar chemicals in the test set (and the similar chemicals v predicted value.



Chemicals	MAE*
Entire set	0.55
Similarity coefficient ≥ 0.5	0.25

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

CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/L)	Predicted value -Log10(mol/L)
91-20-3 (test chemical)			4.32	4.22
90-12-0		0.92	4.20	4.39
98-82-8		0.70	4.28	3.94
106-42-3		0.69	4.10	3.71
141-93-5		0.64	4.51	4.42

Alternatives Assessment Dashboard

- Compare alternatives in terms of color-coded comparison table of L,M, H, VH scores
- Categories include

Acute mammalian toxicity*	Reproductive toxicity	Skin sensitization
Carcinogenicity	Developmental toxicity*	Eye and skin irritation
Mutagenicity*	Neurotoxicity	Acute* and chronic aquatic toxicity
Endocrine disruption*	Systemic toxicity	Persistence and bioaccumulation*

*Endpoints which currently can be predicted by WebTEST

- Data from quantitative toxicity values, GHS scores, QSAR models, government lists

Sources of alternative assessment scores

- Quantitative data
  - US EPA Chemistry Dashboard / RapidTox , Chemidplus
- QSAR Predictions
  - WebTEST
- GHS data
  - Australia, Canada, Denmark, ECHA, Japan, Korea, Malaysia, New Zealand, Germany
- Lists
  - DSL, EPA Mid Atlantic Human Health, Health Canada Priority Substance Lists Carcinogenicity, Health Canada Priority Substance Lists Reproductive, IARC, IRIS, NIOSH Potential Occupational Carcinogens, Prop65, Reach Very High Concern List, Report On Carcinogens, SIN, TEDX, TSCA Work Plan, UMD
- Future work
  - QSAR/read across models based on compiled experimental values
  - Quantitative data from REACH dossiers

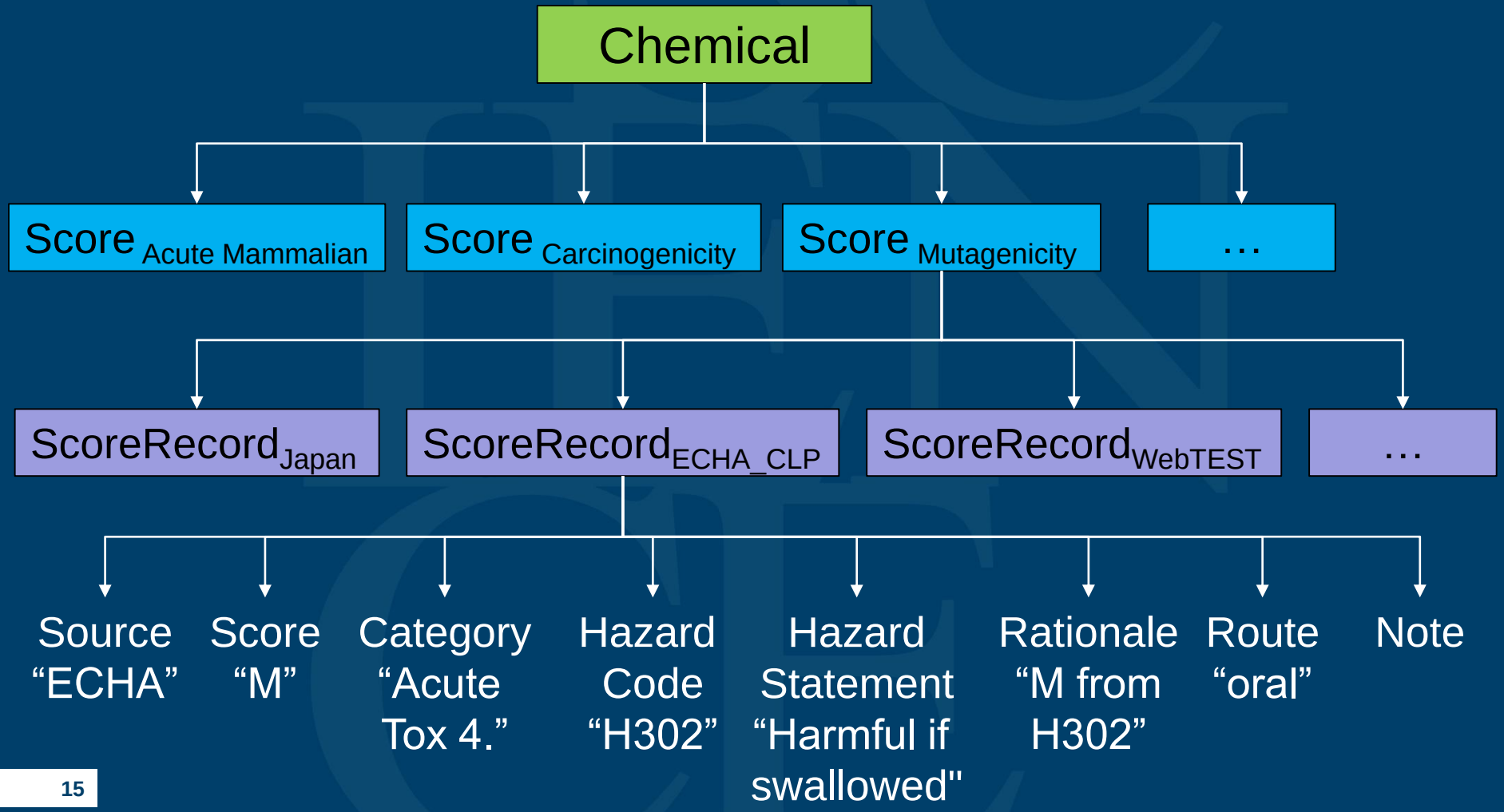
Converting between systems

Acute toxicity scoring

Risk phrase[a]	R28		R25		R22		
LD50, mg/kg	≤5	5-25	25-50	50-200	200-300	300-2000	2000-5000
GHS[b]	Cat 1	Category 2		Category 3		Category 4	Category 5
EU CLP[c]	H300			H301		H302	H303
DfE Score[d]	Very high			High		Moderate	Low

- Quantitative toxicity scores are preferable due to differing systems

AA Data Class Structure



Assigning the final score

- Predicted score is the most toxic score from the source with the highest authority
- Authority levels include Authoritative (e.g. ECHA CLP), Screening (e.g. Chemidplus), and QSAR model (e.g. WebTEST)

Records for Acute Mammalian Toxicity Oral for 56-23-5

Score records

Source	Authority	Score	Route	Category	Hazard Code	Hazard Statement	Rationale
ECHA CLP	Authoritative	H	oral	Acute Tox. 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301
UMD	Screening	VH		Acute toxin			Score of VH was assigned based on a category of Acute toxin
Korea	Screening	H	oral		H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301
New Zealand	Screening	H	oral	Category 6.1C (Category 3)		Acutely toxic	Score was assigned based on a category of Category 6.1C (Category 3).
ChemIDplus	Screening	L	Oral				Oral LD50(5760 mg/kg) > 2000 mg/kg

Alternatives Assessment Dashboard-

- Compare alternatives in terms of color-coded comparison table of L, M, H, VH scores

Comparison of alteratives

Structure CAS name	Human Health Effects															Ecotoxicity		Fate	
	Acute Mammalian Toxicity			Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation
	Oral	Inhalation	Dermal						Repeat Exposure	Single Exposure	Repeat Exposure	Single Exposure							
79-01-6 Trichloroethylene	L	M	L	VH	H	N/A	H	H	H	N/A	H	M	N/A	H	H	H	M	H	L
127-18-4 Perchloroethylene	L	L	L	VH	L	H	M	L	H	H	VH	VH	N/A	VH	H	VH	H	H	L
81-33-4 Pigment Violet 29	M	N/A	N/A	N/A	H	N/A	M	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	L	M	H	L
872-50-4 N – methylpyrrolidone	L	L	L	N/A	L	L	H	H	N/A	N/A	VH	M	L	H	H	L	L	L	L
123-91-1 1,4 - dioxane	M	H	L	VH	H	L	L	L	H	H	VH	M	N/A	H	H	L	L	H	L
106-94-5 1 - Bromopropane	M	M	N/A	VH	L	L	H	H	H	N/A	M	M	L	H	H	M	N/A	H	L
56-23-5 Carbon tetrachloride	H	H	H	VH	L	H	M	L	N/A	H	H	VH	H	H	H	VH	M	H	L
1332-21-4 Asbestos	N/A	N/A	N/A	VH	M	N/A	N/A	N/A	N/A	N/A	VH	N/A	N/A	N/A	N/A	L	N/A	H	L
75-09-2 Methylene chloride	M	L	N/A	VH	H	H	N/A	H	H	H	VH	VH	N/A	H	H	M	L	H	L
25637-99-4 Hexabromocyclododecane	M	N/A	N/A	N/A	H	H	M	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	VH	VH	N/A	L
3194-55-6 1,2,5,6,9,10-Hexabromocyclododecane	L	N/A	N/A	N/A	H	H	M	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	VH	VH	H	H
3194-57-8 1,2,5,6-Tetrabromocyclooctane	L	N/A	N/A	N/A	H	L	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	VH	N/A	N/A	M

AA Comparison Table, cont.

Records for Acute Mammalian Toxicity Oral for 56-23-5

Score records sorted by Authority (descending), Score (descending), and Source (ascending)

Source	Authority	Score	Route	Category	Hazard Code	Hazard Statement	Rationale	Note
ECHA CLP	Authoritative	H	oral	Acute Tox. 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	
UMD	Screening	VH		Acute toxin			Score of VH was assigned based on a category of Acute toxin	Meets the University of Maryland definition of an "Acute Toxin" for the purpose of the University of Maryland Chemical Hygiene Plan. Acute toxins are defined as substances that have a median lethal dose (LD50) less than or equal to 50 mg/kg body weight by the oral route of entry; 200 mg/kg body weight by the dermal route of entry or a median lethal concentration (LC50) less than or equal to 0.5 mg/l where time of exposure is eight hours or less. This definition is compatible with the 1994 Department of Transportation definition of "Poison".
Korea	Screening	H	oral		H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	
New Zealand	Screening	H	oral	Category 6.1C (Category 3)		Acutely toxic	Score was assigned based on a category of Category 6.1C (Category 3).	R-PHRASE: R 25 [N-CLASS]
ChemIDplus	Screening	L	Oral				Oral LD50(5760 mg/kg) > 2000 mg/kg	Test organism: guinea pig Reported Dose: 5760mg/kg Normalized Dose: 5760mg/kg Source: Gigiena i Sanitariya. For English translation, see HYSAAV. Vol. 33(1), Pg. 32, 1968.
ChemIDplus	Screening	L	Oral				Oral LD50(8263 mg/kg) > 2000 mg/kg	Test organism: mouse Reported Dose: 8263mg/kg Normalized Dose: 8263mg/kg Source: Journal of Pharmacy and Pharmacology. Vol. 3, Pg. 169, 1951.
ChemIDplus	Screening	L	Oral				Oral LD50(5760 mg/kg) > 2000	Test organism: rabbit Reported Dose: 5760mg/kg

EPA
United States
Environmental Protection
Agency

Questions???