

Collaborative modeling project for predicting acute oral toxicity (CATMoS)

Kamel Mansouri

LEAD COMPUTATIONAL CHEMIST, ILS IN SUPPORT OF NICEATM

Disclaimer: ILS staff provide technical support for NICEATM, but do not represent NIEHS, NTP, or the official positions of any federal agency.





Collaborative projects

CERAPP

Collaborative Estrogen Receptor
Activity Prediction Project (2015/16)

Mansouri et al. (<https://doi.org/10.1289/ehp.1510267>)

CoMPARA

Collaborative Modeling Project for
Androgen Receptor Activity (2017/18)

Mansouri et al. (<https://doi.org/10.1289/EHP5580>)

CATMoS

Collaborative Acute Toxicity Modeling
Suite (2017/18)

Kleinstreuer et al. (<https://doi.org/10.1016/j.comtox.2018.08.002>)
Mansouri et al. (in preparation)



Endocrine Disruptor Screening Program

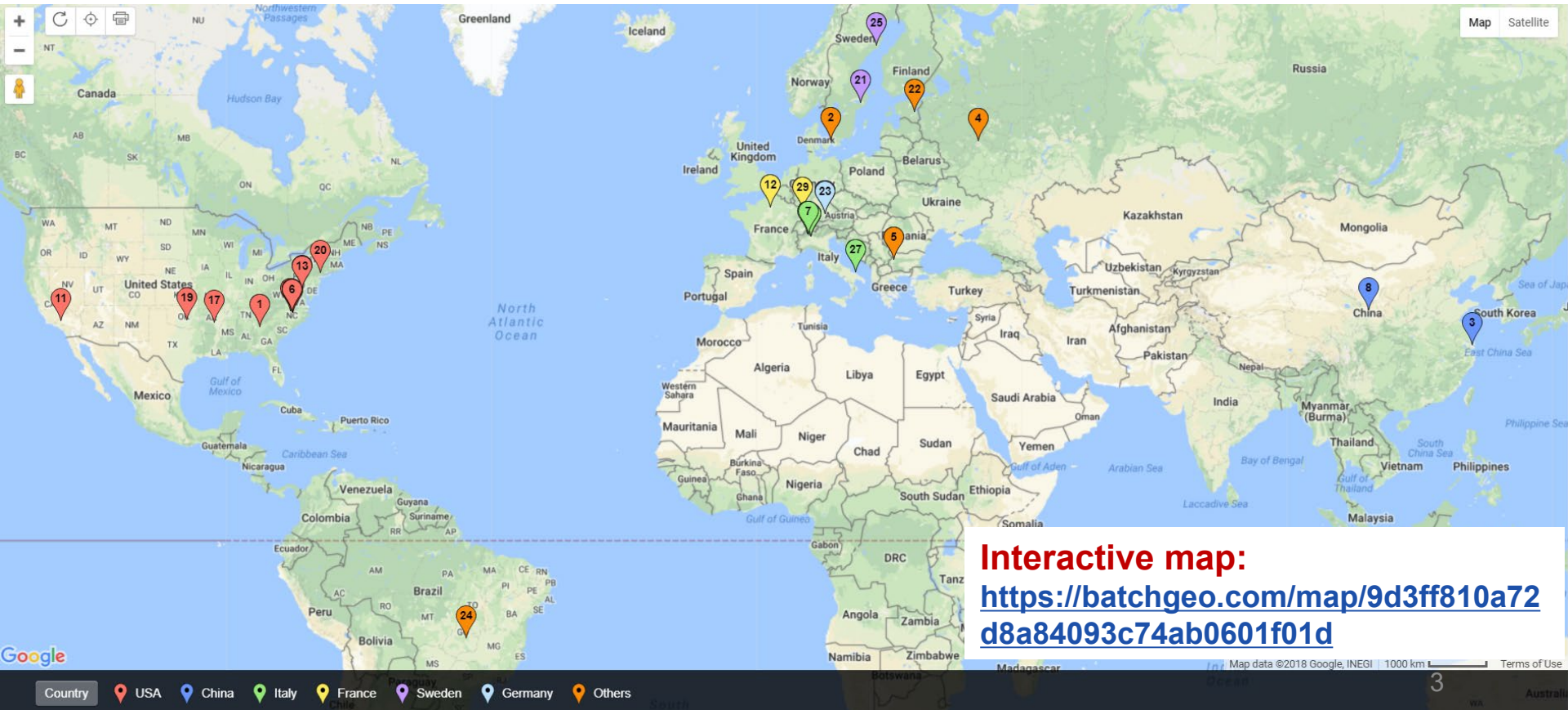


Acute Toxicity Workgroup: alternative methods



International Collaborators

Over 100 scientists from around the globe representing academia, industry, and government contributed





Acute Oral Toxicity: CATMoS

- ICCVAM is developing alternative test methods for the EPA's six pack tests: Acute oral, dermal, inhalation, eye & skin irritation and skin sensitization
- Acute Toxicity Workgroup: identifies federal agency requirements, needs, and decision contexts for using acute systemic toxicity data

Regulatory Toxicology and Pharmacology 94 (2018) 183–196

Contents lists available at [ScienceDirect](#)

 **Regulatory Toxicology and Pharmacology**

journal homepage: www.elsevier.com/locate/yrtph



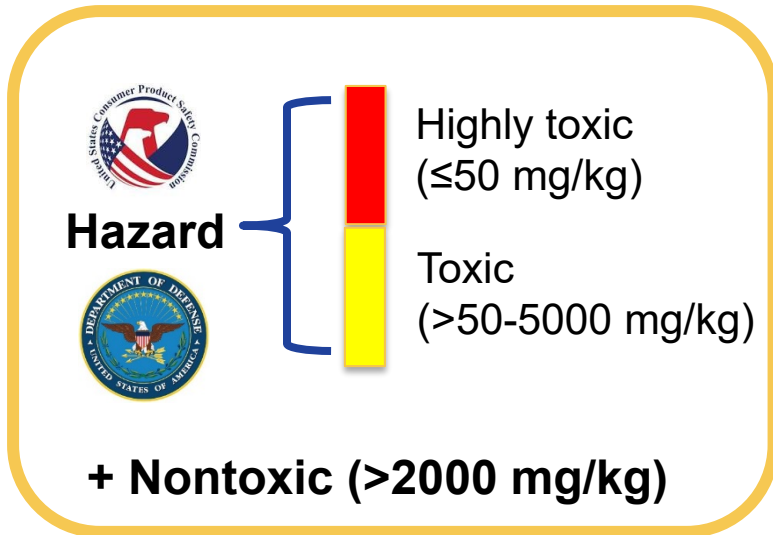
Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies 

Judy Strickland^{a,*}, Amy J. Clippinger^b, Jeffrey Brown^b, David Allen^a, Abigail Jacobs^{c,1}, Joanna Matheson^d, Anna Lowit^e, Emily N. Reinke^f, Mark S. Johnson^f, Michael J. Quinn Jr.^f, David Mattie^g, Suzanne C. Fitzpatrick^h, Surender Ahirⁱ, Nicole Kleinstreuer^j, Warren Casey^j

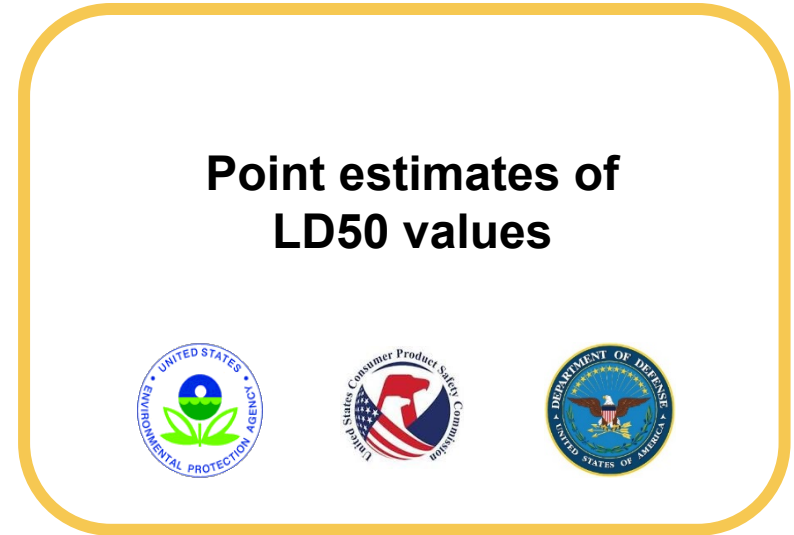


Agency-Based Modeling Endpoint Selection

Binary Models



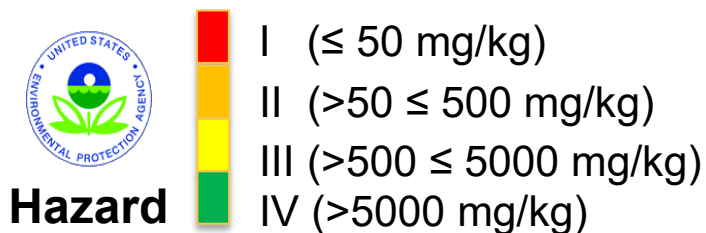
Continuous Model



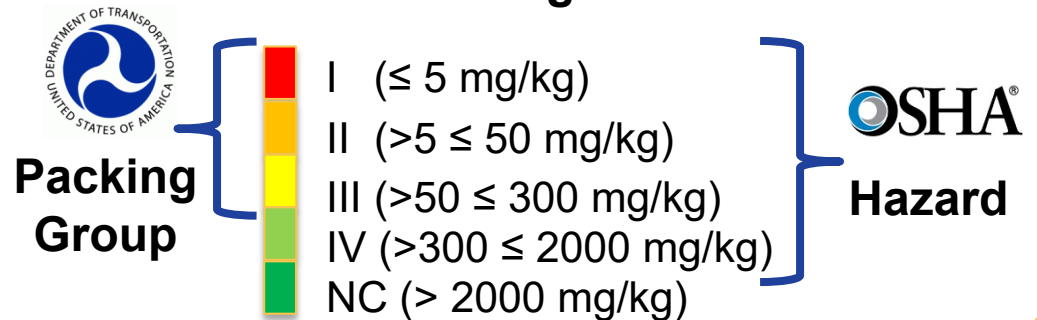
Categorical Models

Hazard

EPA Categories



GHS Categories





Available data for modeling

Rat oral LD50s:
16,297 chemicals total
34,508 LD50 values

15,688 chemicals total
21,200 LD50 values

QSAR-ready standardization

Desalted, stereochemistry stripped,
tautomers and nitro groups standardized,
valence corrected, structures neutralized

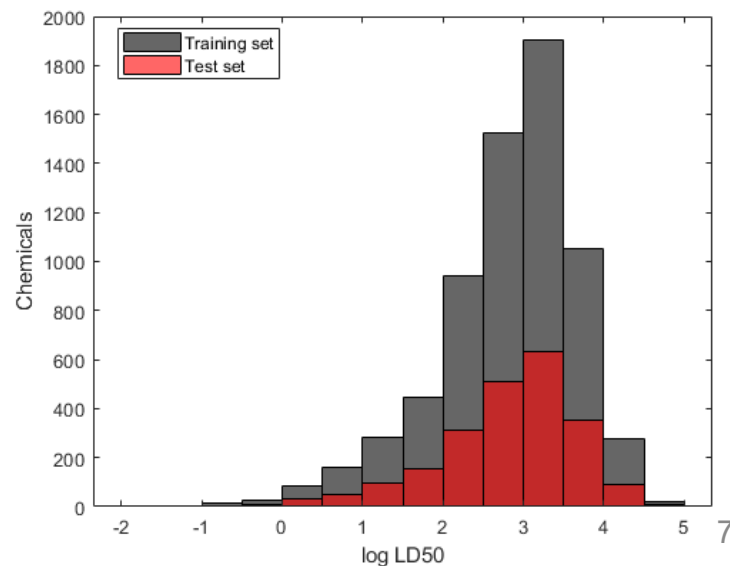
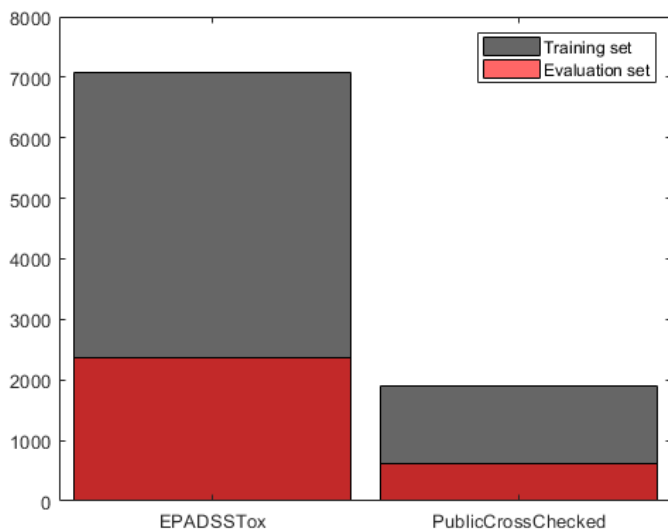
**11992 chemicals with
accurate structures**

- Very toxic endpoint: 11886 entries (binary, 0/1)
- Non-toxic endpoint: 11871 entries (binary, 0/1)
- EPA endpoint: 11755 entries (categorical, 4 categories)
- GHS endpoint: 11845 entries (categorical, 5 categories)
- LD50 endpoint: 8908 entries (continuous values)



Establishing Modeling Dataset

- **Training and evaluation sets:**
 - 11,992 chemicals from the final inventory of chemicals with QSAR-ready structures having rat oral acute toxicity data were split into training and test sets:
 - 75% training set: 8,994 chemicals
 - 25% evaluation set: 2,998 chemicals
 - All endpoints training data included in same structure file
 - Similar distributions and variability for values and categories
 - Similar distribution of chemical structures sources





Establishing Modeling Dataset

- **Prediction set:**

Included lists of regulatory interest:

- ToxCast/Tox21
- EDSP
- TSCA
- Substances on the market (EPA Dashboard list)



After QSAR-ready standardization:

48137 structures to be predicted (including the evaluation set)

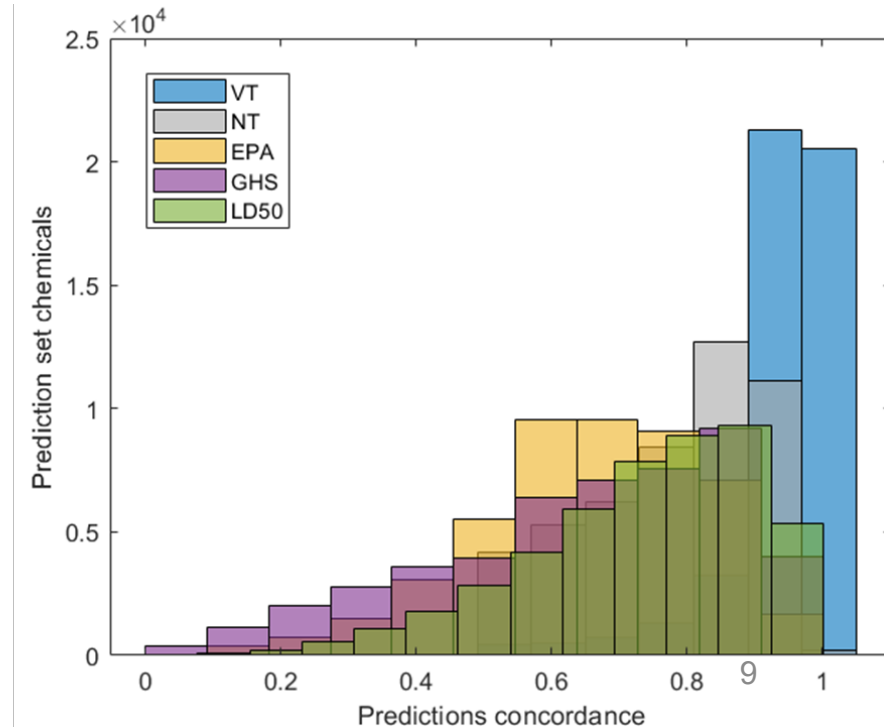
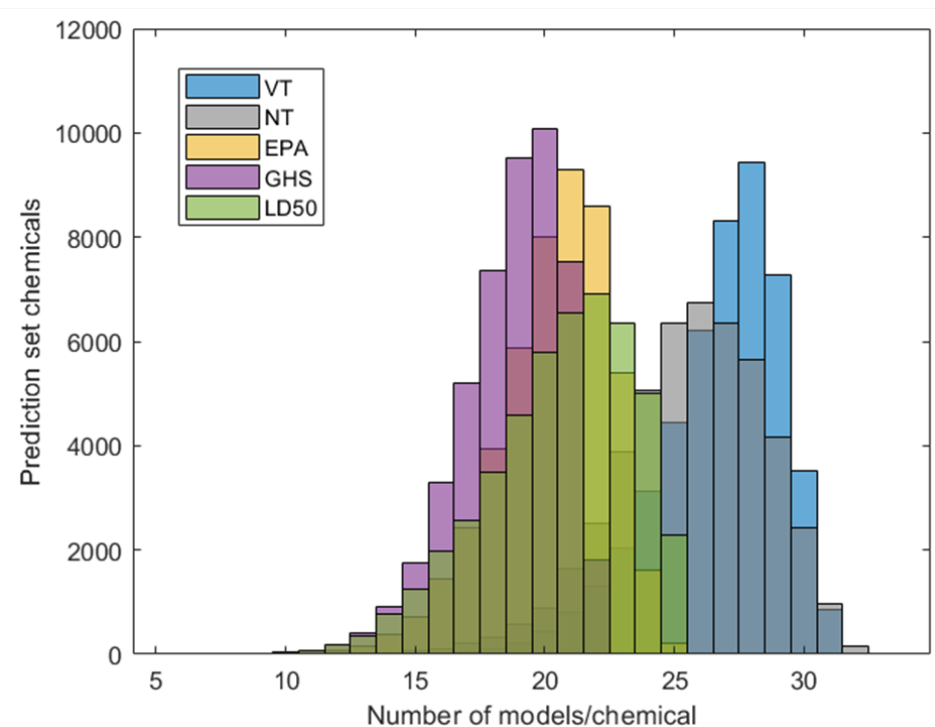


Coverage and concordance of the models

Consortium Comprised 35 Participants/Groups

- Very Toxic: 32 models
- Non-toxic: 33 models
- EPA categories: 26 models
- GHS categories: 23 models
- LD50: 25 models

Total: 139 models





CATMoS consensus modeling

Steps of combining the single models into consensus

Initial models & predictions

- VT (32 models)
- NT (33 models)
- GHS (23 models)
- EPA (26 models)
- LD50 (25 models)

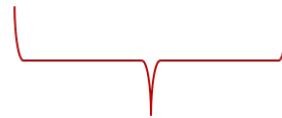
Combining models



Weighted average
/majority rule

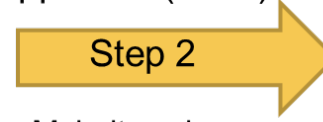
Independent consensus models/predictions

- VT
- NT
- GHS
- EPA
- LD50



A consensus model
per endpoint
(~20-~30 models)

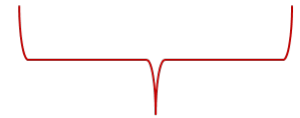
Weight of Evidence
approach (WoE)



Majority rule

Consistent consensus models/predictions

- VT
- NT
- GHS
- EPA
- LD50



Consensus
representing all
~140 models

Learn more:

https://www.piscltd.org.uk/wp-content/uploads/2020/01/2020.01.22_CATMoS_Webinar.pdf

<https://youtu.be/KjbTnfRTY-0>



Performance Assessment

Consensus Model Statistics

	Very Toxic		Non-Toxic		EPA		GHS	
	Train	Eval	Train	Eval	Train	Eval	Train	Eval
Sensitivity	0.87	0.70	0.88	0.67	0.81	0.62	0.80	0.58
Specificity	0.99	0.97	0.97	0.90	0.92	0.86	0.95	0.90
Balanced Accuracy	0.93	0.84	0.92	0.78	0.87	0.74	0.88	0.74
<i>In vivo</i> Balanced Accuracy	0.81		0.89		0.82		0.79	

	LD50 values		LD50 values
	Train	Eval	<i>In Vivo</i>
R2	0.85	0.65	0.80
RMSE	0.30	0.49	0.42

The consensus predictions perform just as well as replicate *in vivo* data do at predicting oral acute toxicity outcome



Collaboration with ATWG partners and ICCVAM agencies

Agency	No. Substances	Agency	No. Substances
Air Force	421	EPA OPP	36
Army Public Health Command	18	EPA OPPT	8
Army Edgewood Chemical Biological Center	42	EPA NCCT	4815
CPSC	110	EPA EFED	160
DOT	3671	FDA CFSAN	22

Progress made with EPA EFED:

- Compared CATMoS predictions to risk assessment data on 160 pesticides registered in the last 25 years.
- Determined overlap and discordance leading to additional curation and improvement of the used data and predictions.



Running CATMoS Consensus and other OPERA models

OPERA standalone application:

- Free, open-source & open-data
- Command line & Graphical user interface
- Single chemical and batch mode
- Multiple platforms (Windows and Linux)
- Embeddable libraries (java, C, C++, Python)
- **New: QSAR-ready standardization**

OPERA models:

- Physicochemical properties
- Environmental fate
- ADME properties
- Toxicity endpoints

Input options:

- Structure IDs (CAS, DTXSID, InChIKey)
- Structure files (SMILES, SDF, Mol)

Links:

<https://github.com/NIEHS/OPERA>

<https://ntp.niehs.nih.gov/go/opera>

<https://doi.org/10.1186/s13321-018-0263-1>

```
OPERA_CL

OPERA models for physchem, environmental fate and tox properties.
Version 2.6 (May 2020)

OPERA is a command line application developed in Matlab providing QSAR
models predictions as well as applicability domain and accuracy assessment.

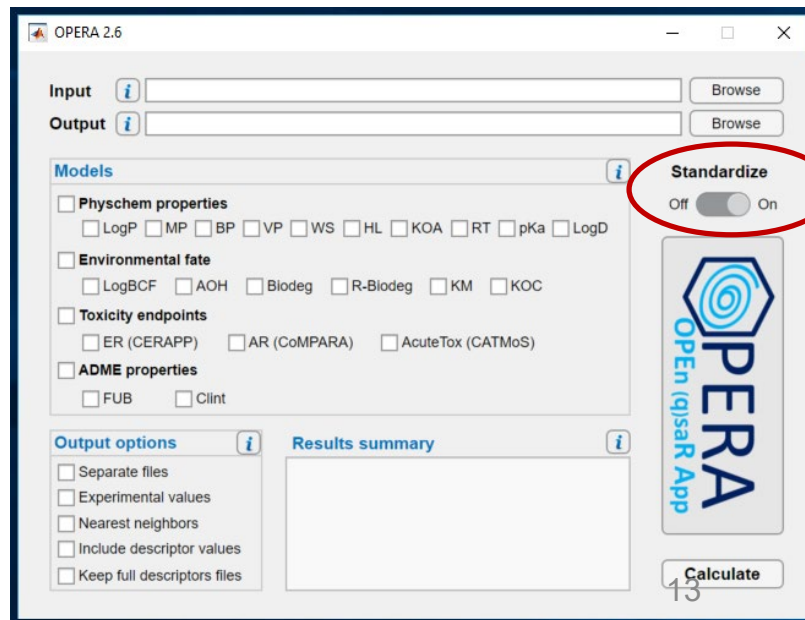
Developed by:
Kamel Mansouri
mansourikamel@gmail.com
kamel.mansouri@nih.gov

Usage: OPERA <argument_list>

Examples:
OPERA -s Sample_50.sdf -o predictions.csv -a -x -v 2
opera -d Sample_50.csv -o predictions.txt -e logP BCF -n -v 1

Type OPERA -h or OPERA --help for more info.

C:\Users\kmansouri>
```





New in OPERA version 2.6

QSAR-ready standardization tool

```
OPERA_CL
C:\Users\kmansouri\Downloads>opera -s PestActivesEPA.sm --standardize -o pred.csv -tox -v 1

Endpoints to be calculated:
CERAPP, COMPARA and CATMOS

Initializing and loading models...

===== Structures standardization =====
Input structures: 510.
Generating QSAR-ready structures...
Standardized structures: 464.

===== Molecular Descriptors =====
Loaded structures: 464.
PaDEL calculating 2D descriptors...
PaDEL descriptors calculated for: 464 molecules.
Loading of PaDEL descriptors file...
Checking loaded variables.
Loaded 1444 PaDEL descriptors for 464 molecules.
CDK 2.0 calculating 2D descriptors...
CDK descriptors calculated for: 464 molecules.
Loading of CDK descriptors file...
Checking loaded variables.
Loaded 286 CDK descriptors for 464 molecules.

===== Running The Models =====
----- Toxicity Endpoints -----
Predicting Estrogen Receptor Activity (CERAPP)...
Predicting Androgen Receptor Activity (CoMPARA)...
Predicting Acute Oral Tox. endpoints (CATMoS)...

===== End Of Calculation =====
464 molecules predicted. Total process time: 00:05:44.

C:\Users\kmansouri\Downloads>
```

Command line

OPERA 2.6

Input *i* C:\Users\kmansouri\Downloads\PestActivesEPA.smi

Output *i* C:\Users\kmansouri\Downloads\PestActivesEPA-smi_OPERA2.6Pred.csv

Models *i*

Physchem properties
 LogP MP BP VP WS HL KOA RT pKa LogD

Environmental fate
 LogBCF AOH Biodeg R-Biodeg KM KOC

Toxicity endpoints
 ER (CERAPP) AR (CoMPARA) AcuteTox (CATMoS)

ADME properties
 FUB ClInt

Output options *i*

Separate files
 Experimental values
 Nearest neighbors
 Include descriptor values
 Keep full descriptors files

Results summary *i*

Loaded structures from SMILES file: 510
Standardized structures: 464
Calculated PaDEL descriptors: 1444 (119 sec)
Calculated CDK descriptors: 286 (58 sec)
Predicted structures: 464 (233 sec)
Total processing time: 00:08:14.

Standardize
Off On

OPERA
Open (q)SAR App

Graphical user interface



Example output

SMI SMILES	S MoleculeID	I CATMoS_VT_pred	I CATMoS_NT_pred	I CATMoS_EPA_pred	I CATMoS_GHS_pred	S CATMoS_LD50_exp	D CATMoS_LD50_pred	S CATMoS_LD50_predRange	I AD_CATMoS	D AD_index_CATMoS	D Conf_index_CATMoS
	104-55-2	0	1	3	5	2220	2,568	[1300-5100]	1	1	0.958
	78-70-6	0	1	3	5	2795	2,218	[1100-4400]	1	1	0.958
	103-90-2	0	0	3	4	501-5000	1,625	[810-3200]	1	1	0.964
	107-02-8	1	0	1	2	20	40	[20-80]	1	1	0.772
	115-29-7	1	0	1	1	NA	2.26	[1-4,5]	1	1	0.823

- Consensus predictions for the 5 endpoints
- LD50 confidence interval (based on in vivo data variability)
- Applicability domain assessment
- Experimental values, when available
- Nearest neighbors, optional



Predictions on NTP/ICE

Chemical Input

Select Chemicals

Quick List CASRN

User CASRNs

- 104-55-2
- 78-70-6
- 103-90-2
- 107-02-8
- 115-29-7

Add chemicals with identical QSAR structures

Assay Input

Select Assays

Assay

- CERAPP, ER Binding
- CERAPP, ER Antagonist
- CERAPP, ER Agonist
- CoMPARA, AR Binding
- CoMPARA, AR Antagonist
- CoMPARA, AR Agonist
- CATMoS, Rat Acute Oral Toxicity

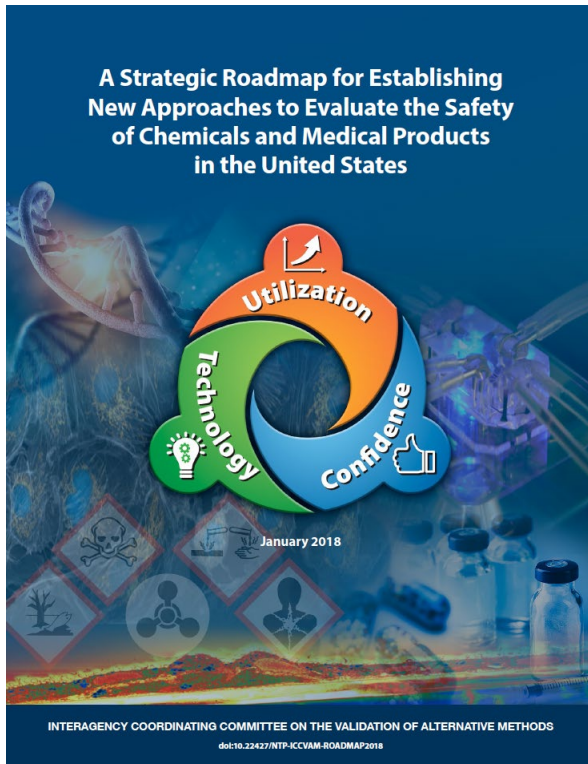
<https://ice.ntp.niehs.nih.gov/Search>

Download	Query Mixtures	Clear Filter	Number of chemicals = 5						
Substance Name	CASRN	DTXSID	CATMoS, Rat Acute Oral Toxicity LDS0	CoMPARA, AR Agonist Call	CoMPARA, AR Antagonist Call	CoMPARA, AR Binding Call	CERAPP, ER Agonist Call	CERAPP, ER Antagonist Call	CERAPP, ER Binding Call
Acetaminophen	103-90-2	DTXSID2020006	1625	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
Endosulfan	115-29-7	DTXSID1020560	2.26	Inactive	Inactive	Inactive	Inactive	Active	Active
3-Phenylprop-2-enal	104-55-2	DTXSID1024835	2568	Inactive	Inactive	Inactive	Inactive	Active	Active
Acrolein	107-02-8	DTXSID5020023	40	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
Linalool	78-70-6	DTXSID7025502	2097	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive



The “3C” Concept at Work!

- Success of the projects was due in great part to the use of the 3C concept as well as up-front and continuous engagement of regulators in the process



Communication



Collaboration



Commitment



Thank you for your attention!

Acknowledgements

- EPA/CCTE
- ICCVAM Acute Toxicity Workgroup
- NTP/NICEATM
- ILS
- ScitoVation
- All international collaborators

Funding

- ILS/NICEATM under NIEHS contract HHSN273201500010C
- EPA/ORD, Oak Ridge Institutes through U.S. DoE & EPA.
- The Lush Prize 2017, young researchers, supporting animal free testing.