Collaborative Modeling Project for Predicting Acute Oral Toxicity

Kamel Mansouri[1], Agnes Karmaus[1], Jeremy Fitzpatrick[2,3], Grace Patlewicz[2], Prachi Pradeep[2], David Allen[1], Warren Casey[4], and Nicole Kleinstreuer[4]

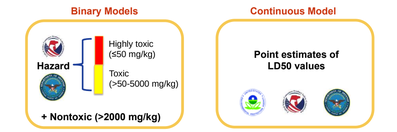
[1]ILS, RTP, NC, USA; [2]EPA/CCTE, RTP, NC, USA; [3]ScitoVation LLC, RTP, NC, USA; [4]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

Background

- Acute systemic toxicity tests are commonly required by regulatory authorities to characterize a chemical's toxicity.
- In silico models provide an alternative to traditional animal tests for predicting acute oral toxicity and bridging data gaps.
- NICEATM and the ICCVAM Acute Toxicity Workgroup (ATWG) organized an international collaborative project to develop in silico models for predicting acute oral toxicity.
- Predictions within the applicability domains of the submitted models were evaluated using external validation sets, then combined into consensus predictions for each endpoint, forming the Collaborative Acute Toxicity Modeling Suite (CATMoS).

Project Data

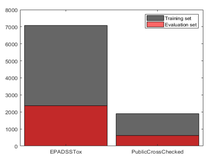
- Endpoints: five endpoints were selected by the ICCVAM ATWG member agencies to serve as endpoints for predictive modeling within the CATMoS project.
- Collected data: 34,508 rat oral LD50 values for 16,297 chemicals total.

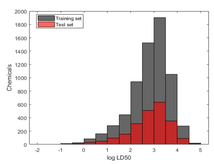


15,688 chemical structures
21,200 LD50 values

QSAR-ready standardization
Desalted, stereochemistry stripped, tautomers and nitro groups standardized, valence corrected, structures neutralized

11,992 chemicals with standardized structures





Available data split into:
- 75% training set: 8,994 chemicals
- 25% evaluation set: 2,998 chemicals

- Training data for all endpoints included in same structure file
- Similar distributions and variability for values and categories
- Similar distribution of chemical structure sources

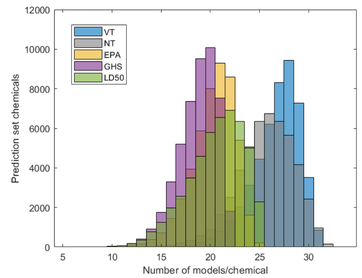
International Consortium of Participants

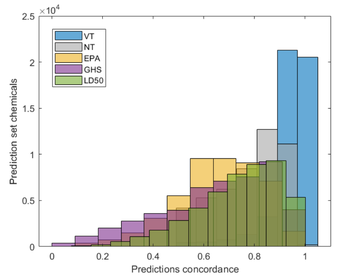
A consortium of **35 international participants** representing academia, industry, and government

Group ID	Institution	Country
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods	USA
UNIBARI	Università degli Studi di Bari	Italy
LOREAL	L'Oréal R&I	France
UNICAMB	University of Cambridge	UK
UNC	UNC Eshelman School of Pharmacy	USA
FUG	Federal University of Goias	Brazil
UNIMIB	University of Milano-Bicocca	Italy
DOW	The Dow Chemical Company	USA
IRCCS (5 groups)	Istituto di Ricerche Farmacologiche Mario Negri	Italy
MSU	Michigan State University	USA
SIMPLUS	Simulations Plus, Inc.	USA
KU	Kyoto University Graduate School of Medicine	Japan
ECUST	East China University of Science and Technology, China	China
USAFSAM	Henry M Jackson Foundation for the Advancement of Military Medicine	USA
RUT (2 groups)	Rutgers University	USA
COLPHA	Collaborations Pharmaceuticals, Inc.	USA
UL	Underwriters Laboratories	USA
NCSTATE	North Carolina State University	USA
PNNL	Pacific Northwest National Laboratory	USA
NCCT	National Center for Computational Toxicology, USEPA	USA
HZM	Helmholtz Zentrum München, Germany	Germany
UNISTRA	Universite de Strasbourg	France
NRMRL	National Risk Management Research Laboratory, USEPA	USA
LSINC	Leadscope Inc.	USA
NCATS	National Center for Advancing Translational Sciences, NIH	USA
ATSDR	Agency for Toxic Substances and Disease Registry, CDC	USA
ROSETTAC	Rosettastein Consulting UG	Germany
UCOL	University of Colorado	USA
DUT	Dalian University of Technology	China
DOW_AGRO	Dow Agrosciences	USA

Consensus Modeling

Coverage and concordance of the models (139 models received)





Model evaluation procedure

Qualitative evaluation:
- Documentation
- Defined endpoint
- Unambiguous algorithm
- Availability of code
- Defined applicability domain
- Availability of input data used for modeling
- Mechanistic interpretation

Quantitative evaluation:
- Goodness of fit: training (Tr) statistics
- Predictivity: statistics on the evaluation set (Eval)
- Robustness: balance between (Goodness of fit) & (Predictivity)

$$S = 0.3 * (Goodness\ of\ fit) + 0.45 * (Predictivity) + 0.25 * (Robustness)$$

Categorical models (binary and multi-class):

$$Goodness\ of\ fit = 0.7 * (BA_{Tr}) + 0.3 * \left(1 - |Sn_{Tr} \widetilde{-} Sp_{Tr}|\right)$$
$$Predictivity = 0.7 * (BA_{Eval}) + 0.3 * \left(1 - |Sn_{Eval} \widetilde{-} Sp_{Eval}|\right)$$
$$Robustness = 1 - |BA_{Tr} - BA_{Eval}|$$

$$BA = \frac{(Sn + Sp)}{2} \quad Sn = \frac{TP}{TP + FN} \quad Sp = \frac{TN}{TN + FP}$$

Continuous models:

$$Goodness\ of\ fit = R^2_{Tr}$$
$$Predictivity = R^2_{Eval}$$
$$Robustness = 1 - |R^2_{Tr} - R^2_{Eval}|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{TR}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{TR}} (y_i - \bar{y})^2} \quad \hat{y}_i \text{ and } y_i \text{ are the estimated and observed responses}$$

Steps for combining the models into consensus

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

Combining models (per endpoint)
Step 1
Weighted average /majority rule

Independent consensus models/predictions
- VT
- NT
- GHS
- EPA
- LD50

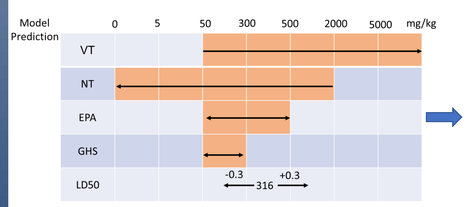
A consensus model per endpoint (~20-~30 models)

Weight of Evidence approach (**WoE**)
Step 2
Majority rule

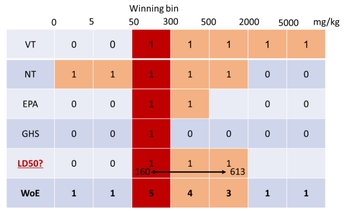
Consistent consensus models/predictions
- VT
- NT
- GHS
- EPA
- LD50

Consensus representing all **139** models

WoE approach to combine the five independent calls



Variability range (log units) for LD50



Adjusted LD50: (160+300)/2=230mg/kg

Original: independent calls

	VT	NT	EPA	GHS	LD50
molX	0	0	2	3	316

WoE: consistent calls

	VT	NT	EPA	GHS	LD50
molX	0	0	2	3	230

CATMoS Performance Evaluation

	LD50	
	Training	Evaluation
R^2	0.85	0.65
RMSE	0.30	0.49

	VT		NT	
	Training	Evaluation	Train	Evaluation
Balanced accuracy (BA)	0.93	0.84	0.92	0.78
Sensitivity (Sn)	0.87	0.70	0.88	0.67
Specificity (Sp)	0.99	0.97	0.97	0.90

	EPA Training				EPA Evaluation			
	Cat 1	Cat 2	Cat 3	Cat 4	Cat 1	Cat 2	Cat 3	Cat 4
BA	0.87				0.74			
Sn	0.87	0.83	0.91	0.63	0.70	0.56	0.81	0.40
Sp	0.99	0.95	0.75	0.98	0.97	0.88	0.62	0.97

	GHS Training					GHS Evaluation				
	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
BA	0.88					0.74				
Sn	0.73	0.75	0.84	0.80	0.88	0.50	0.53	0.56	0.66	0.67
Sp	0.99	0.99	0.92	0.89	0.96	0.99	0.97	0.89	0.74	0.90

CATMoS in Practice

Extended consensus model using a weighted read-across approach



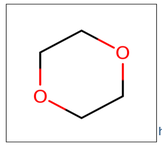
$d_1 = 0$
$Pred_i = N_i$



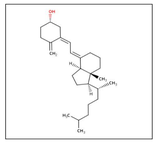
$d_1 \neq 0$
$w_i = f(d_i)$
$Pred_i = f(w_i, N_i)$

★ New chemical to be predicted
● Nearest neighbors (N_i)
d_i: Euclidean distance based on the selected descriptors for each endpoint
➡ Automated, similarity-endpoint dependent read-across: weighted kNN

Example predictions. Structures taken from the EPA CompTox Chemicals Dashboard.



1,4-Dioxane
123-91-1 | DTXSID4020533
Molecular Formula: $C_4H_8O_2$
Average Mass: 88.106 g/mol
LD50: 4200 mg/kg
https://comptox.epa.gov/dashboard/



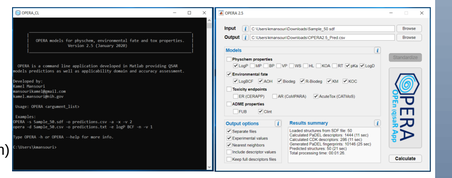
Vitamin D3
67-97-0 | DTXSID6026294
Molecular Formula: $C_{27}H_{44}O$
Average Mass: 384.648 g/mol
LD50: 42 mg/kg
https://comptox.epa.gov/dashboard/

Consensus output: Exported results sheet with predictions, confidence range, applicability domain, and accuracy estimates.

MoleculeID	CATMoS_VT_pred	CATMoS_NT_pred	CATMoS_EPA_pred	CATMoS_GHS_pred	CATMoS_LD50_pred	CATMoS_LD50_predRange	AD_CATMoS	AD_index_CATMoS	Conf_index_CATMoS
123-91-1	0	1	3	5	2543	[1300-5100]	1	1	0.95
67-97-0	1	0	1	2	19	[10-38]	1	1	0.868421053

CATMoS implementation in OPERA

OPERA suite of models:
- Free, open-source, and open-data
- Command line and GUI
- Single chemical and batch mode
- Windows OS and Linux
- Embeddable wrapper libraries (java, C/C++, Python)



References

- Strickland et al. 2018. Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. Regul Toxicol Pharmacol 94:183–196.
- Predictive Models for Acute Oral Systemic Toxicity. https://ntp.niehs.nih.gov/go/tox-models. Accessed 29 Aug 2019.
- Kleinstreuer et al. 2018. Predictive models for acute oral systemic toxicity: A workshop to bridge the gap from research to regulation. Comput Toxicol 8:21–24.
- OPERA – Open structure-activity/property relationship app. https://github.com/NIEHS/OPERA. Accessed 29 Aug 2019.
- Mansouri et al. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform 10:10.
- Integrated Chemical Environment (ICE). https://ice.ntp.niehs.nih.gov/. Accessed 29 Aug 2019.
- Williams et al. 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9(1):61. https://comptox.epa.gov/dashboard/

Acknowledgements

This project was funded in whole or in part with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN273201500010C.

Disclaimer: The views expressed above do not necessarily represent the official positions of any federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.