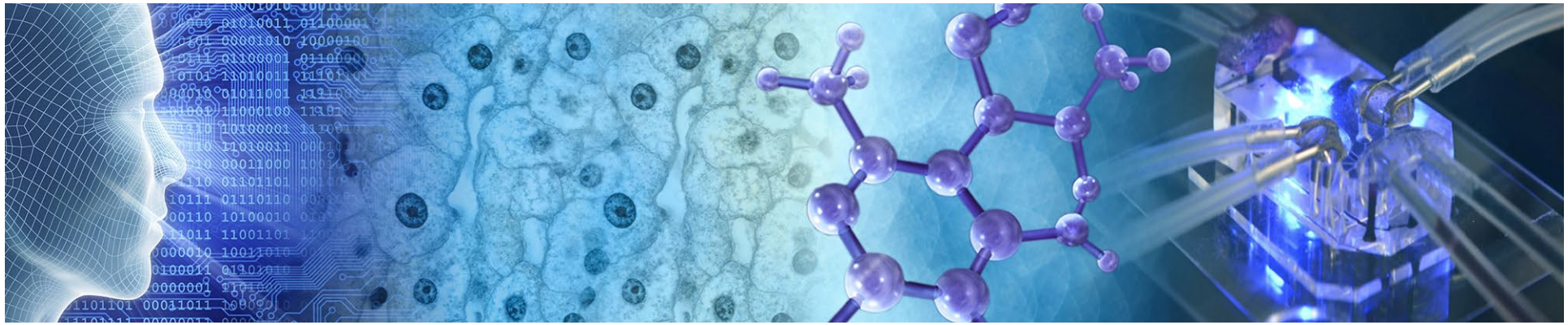
National Institute of
Environmental Health Sciences
Division of Translational Toxicology

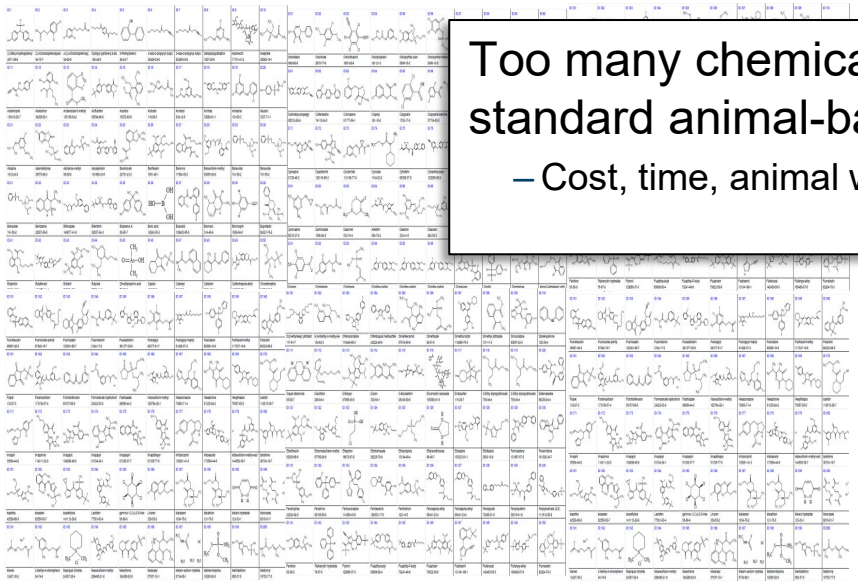


OPERA: Open-Source QSAR Models for Regulatory Support

Kamel Mansouri
Computational Chemist
NIH/NIEHS/DTT/NICEATM, RTP, NC, USA

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of any federal agency.

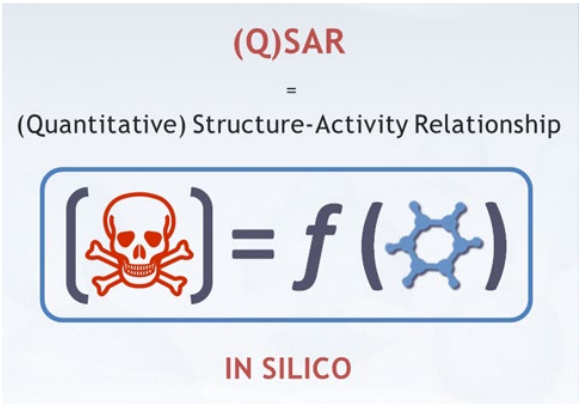
In silico screening



Too many chemicals to test with standard animal-based methods

– Cost, time, animal welfare

Alternative

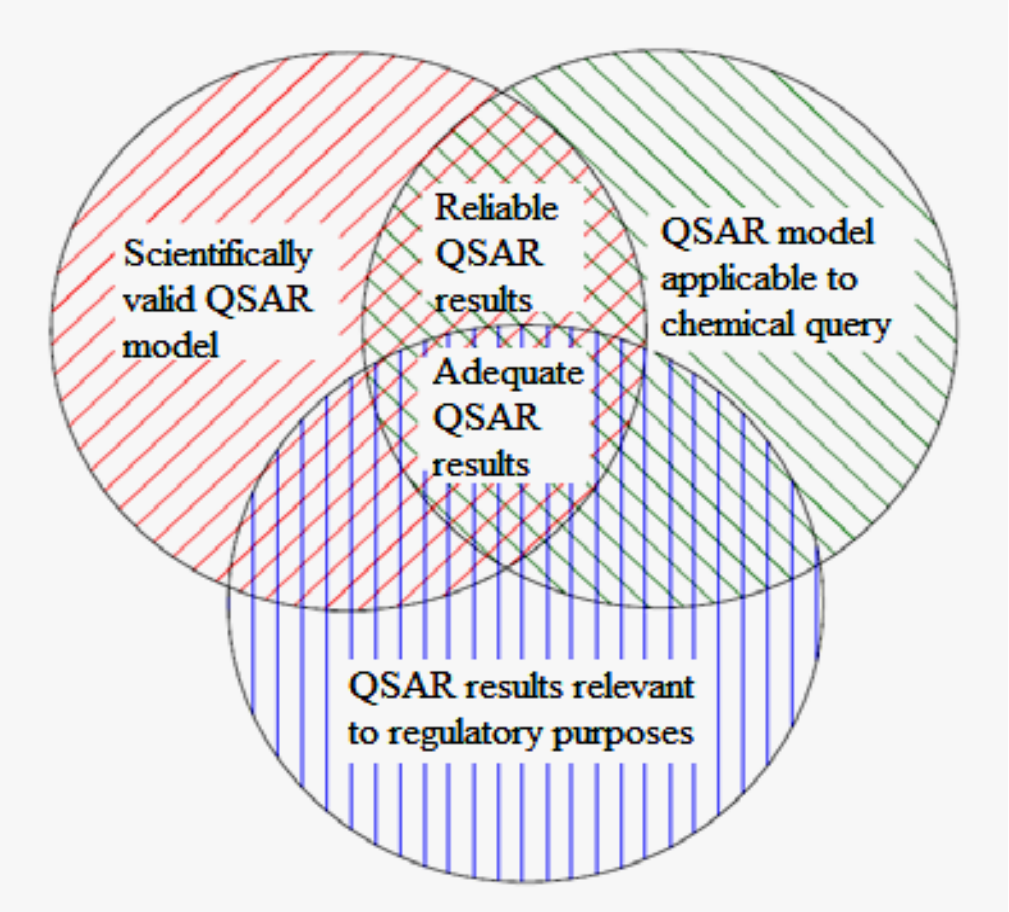


- Organic **pollutants** with exposure potential **accumulate** in body tissues
  - ➢ Cause **toxic effects** to wildlife and humans
- Existence of **gaps in the experimental data** for environmental endpoints
  - ➢ Need to fill the data gaps and bridge the **lack of knowledge**
- **Regulatory** requirements:
  - ➢ Reduce **animal** testing**, time** and **costs**

➢ **Methodology:** use of **QSAR/QSPR** to **predict** the **endpoints** of interest**.**

QSARs for regulatory purposes

The 5 OECD Principles
1) A defined endpoint
2) An unambiguous algorithm
3) A defined domain of applicability
4) Appropriate measures of goodness-of-fit, robustness and predictivity
5) Mechanistic interpretation, if possible



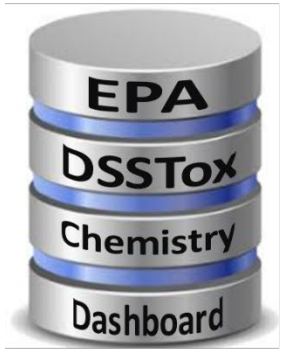
Journal of Cheminformatics
December 2018, 10:10 | Cite as

OPERA models for predicting physicochemical properties and environmental fate endpoints

OPERA
OPEn (q)saR App

OPERA approach

- Curated **open** access datasets (https://doi.org/10.1186/s13321-018-0263-1)
- **Open-source** code (github.com/NIEHS/OPERA)
- **Transparent** unambiguous algorithms (https://qsardb.jrc.ec.europa.eu/qmrf/)
- **Transparent** validated performances (https://doi.org/10.1080/1062936X.2016.1253611)
- **Defined** applicability domain and limitations of the models
- Predictions **available** through:
  - NICEATM's Integrated Chemical Environment (https://ice.ntp.niehs.nih.gov/)
  - The EPA's CompTox Dashboard (https://comptox.epa.gov/dashboard)
  - Free and open-source standalone application (github.com/NIEHS/OPERA)



Integrated Chemical Environment

National Institute of Environmental Health Sciences
Division of Translational Toxicology

OPERA as a standalone desktop application

OPERA standalone application:

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

OPERA models:

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

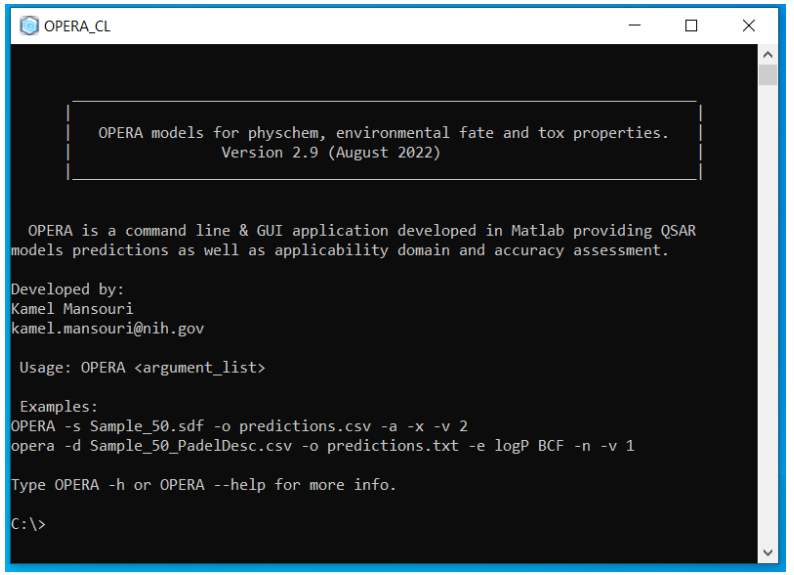
Input options:

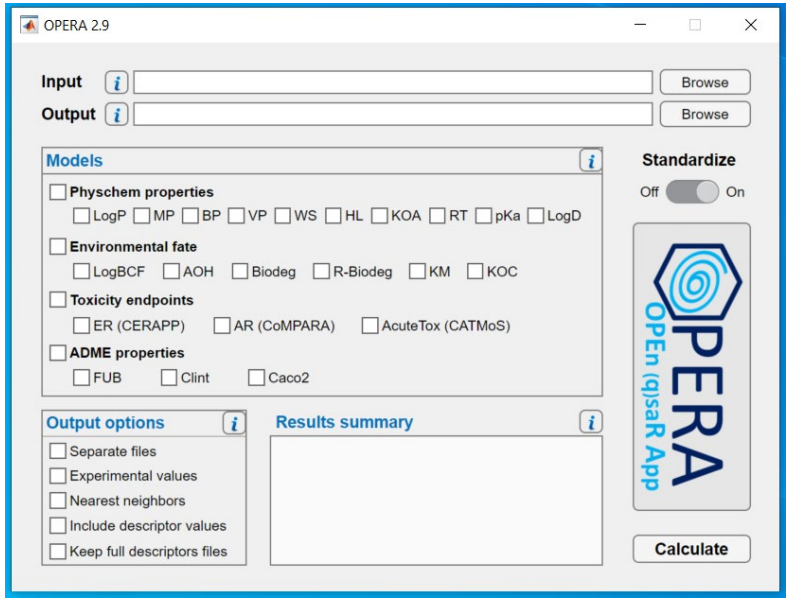
- Structure IDs (CAS, DTXSID, InChIKey)
- Structure files (SMILES, SDF, Mol)

Download and learn more:

https://github.com/NIEHS/OPERA

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





National Institute of Environmental Health Sciences
Division of Translational Toxicology

OPERA models (version 2.9)

Physchem properties		Chemicals	Version
BP	**Boiling Point**	**7860**	**2.9**
HL	**Henry's Law Constant**	**2233**	**2.9**
LogP	**Octanol-water Partition Coefficient**	**18154**	**2.9**
MP	**Melting Point**	**22554**	**2.9**
VP	**Vapor Pressure**	**6764**	**2.9**
WS	**Water Solubility**	**9943**	**2.9**
pKa	Acid Dissociation Constant	6503	2.6
KOA	Octanol/Air Partition Coefficient	270	2.6

Environmental fate		Chemicals	Version
AOH	Atmospheric Hydroxylation Rate	692	2.6
BCF	Bioconcentration Factor	626	2.6
BioHL	Biodegradation Half-life	150	2.6
RB	Ready Biodegradability	1603	2.6
KM	Fish Biotransformation Half-life	541	2.6
KOC	Soil Adsorption Coefficient	728	2.6

Toxicity endpoints		Chemicals	Version
ER	Estrogen Receptor Activity	32464	2.6
AR	Androgen Receptor Activity	47673	2.6
AcuteTox	Acute Oral Systemic Toxicity	50660	2.6

ADME properties		Chemicals	Version
FUB	**Fraction unbound**	**3229**	**2.8**
Clint	**Intrinsic clearance**	**1346**	**2.8**
CACO2	**Caco-2 permeability**	**4601**	**2.8**

New/updated since 2021

Version 3 coming soon!

PFAS in OPERA models

	Models	PFAS/Total
BP	Boiling Point	346/7860
LogP	Octanol-water Partition Coef.	97/18154
MP	Melting Point	409/22554
VP	Vapor Pressure	178/6764
WS	Water Solubility	105/9943

VP Vapor Pressure

Deviation of log^{V}p (Pa) estimates from experimental measurements at 25 °C

	COSMOtherm	EPI Suite	NICEATM	ACD/Labs	TEST	OPERA
MAE	1.22	1.08	1.46	1.53	1.31	0.95
RMSE	1.48	1.48	2.06	1.99	1.91	1.26

LogP Octanol-water Coef

Deviation of octanol–water partition ratio (log K_{OW}) estimates from experimental measurements at 25 °C

	COSMOtherm	EPI Suite	NICEATM	ACD/Labs	OPERA	LSER
MAE	0.41	0.25	0.68	0.61	0.21	0.33
RMSE	0.50	0.29	1.13	0.70	0.28	0.36

WS Water solubility

Deviation of water solubility (log S; mg/L) estimates from experimental measurements at 25 °C

	COSMOtherm	EPI Suite	NICEATM	TEST[a]	OPERA
MAE	0.35	1.82	2.38	0.95	0.23
RMSE	0.41	2.20	2.55	1.36	0.36

Environmental Chemistry

Property Estimation of Per- and Polyfluoroalkyl Substances: A Comparative Assessment of Estimation Methods

Alina Lampic and J. Mark Parnis*

Chemical Properties Research Group (Canadian Environmental Modelling Centre), Department of Chemistry, Trent University, Peterborough, Ontario, Canada

International 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)

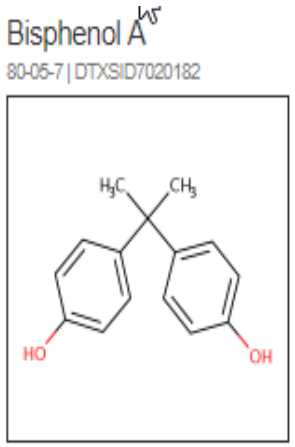
Kleinstreuer et al. (https://doi.org/10.1016/j.comtox.2018.08.002)
Mansouri et al. (https://doi.org/10.1289/EHP8495)

Endocrine Disruptor Screening Program

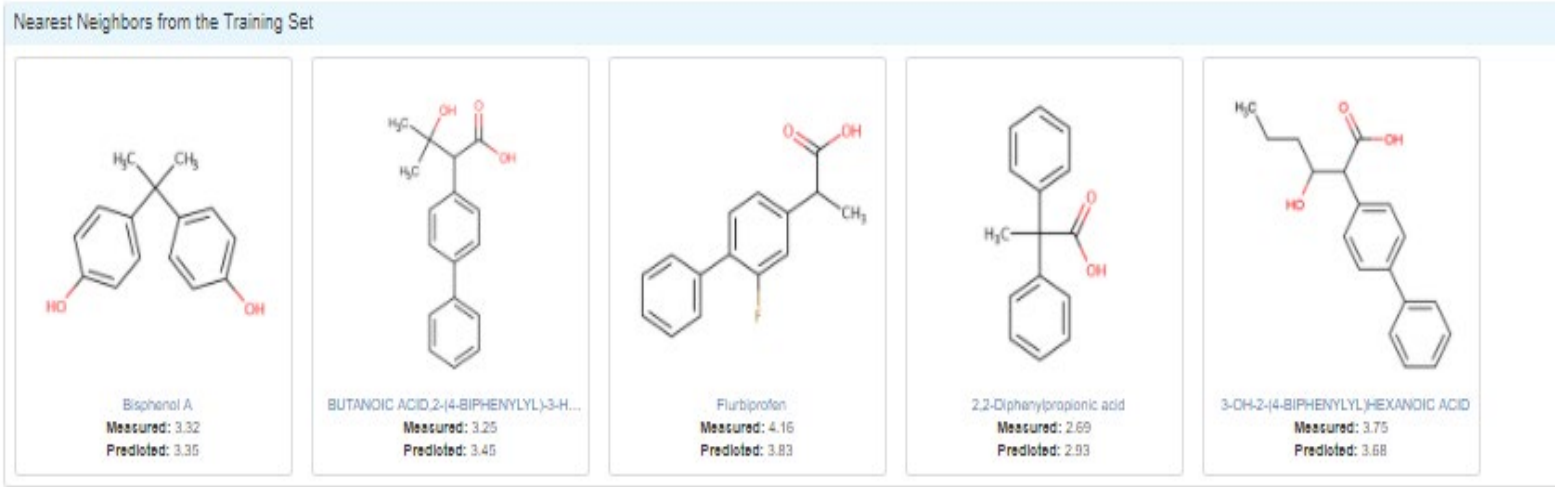
Acute Toxicity Workgroup: alternative methods

ICCVAM: Interagency Coordinating Committee on the Validation of Alternative Methods

OPERA output



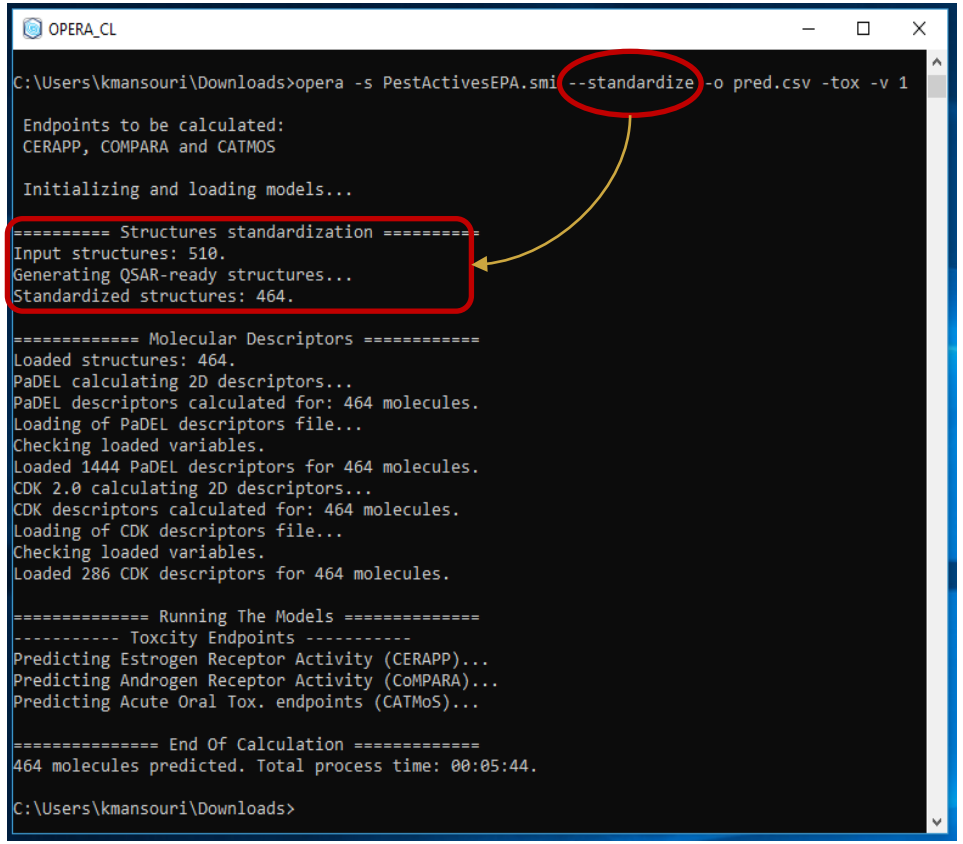
- Predicted value
- Experimental value, if available
- Global applicability domain
- Local applicability domain index [0:1]
- Accuracy index [0:1]
- Confidence interval [min:max]
- 5 nearest neighbors:
  - Identifiers (CAS, DTXSID, InChiKey)
  - Experimental values
  - Predicted values
- Molecular descriptors values

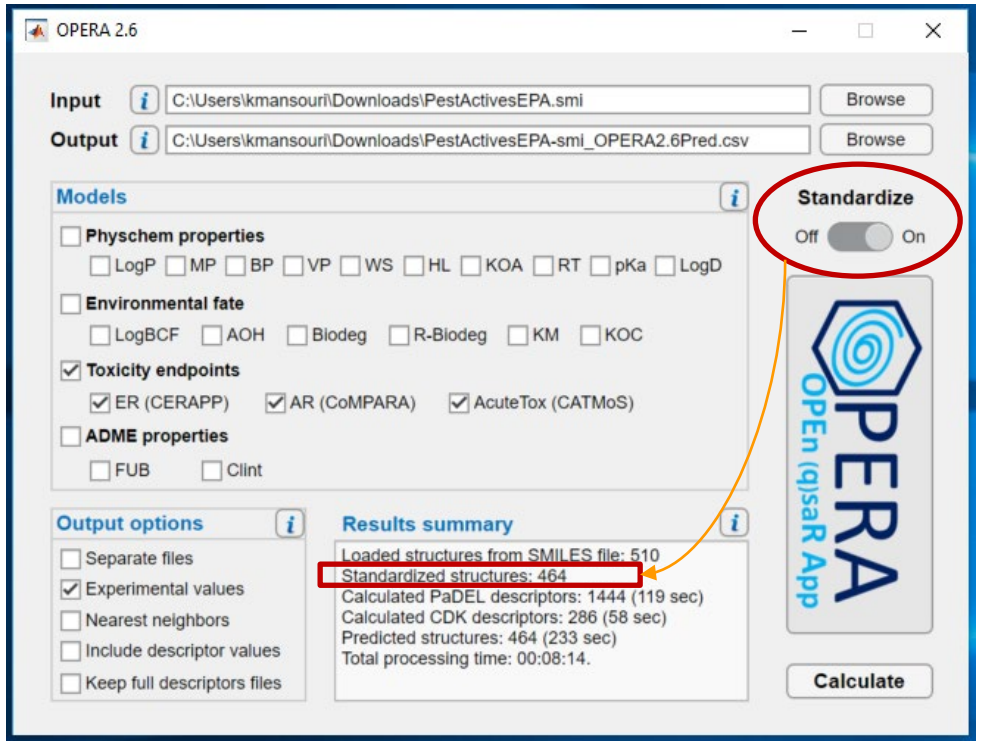


	MoleculeID	DTXSID7020182
	LogP_exp	3.32
	LogP_pred	3.32
	LogP_predRange	[3.27:3.37]
	AD_LogP	1
	AD_index_LogP	1
	Conf_index_LogP	0.948
	LogP_CAS_neighbor_1	80-05-7
	LogP_CAS_neighbor_2	93371-55-2
0	LogP_CAS_neighbor_3	85-27-8
1	LogP_CAS_neighbor_4	5104-49-4
2	LogP_CAS_neighbor_5	5558-66-7
3	LogP_InChiKey_neighbor_1	IISBACLAFKSPIT-U
4	LogP_InChiKey_neighbor_2	ZNPAJPOVURPJFI
5	LogP_InChiKey_neighbor_3	PQSXNIMHIHYFE
5	LogP_InChiKey_neighbor_4	SYTBZMRGLBWN
7	LogP_InChiKey_neighbor_5	ODELFXJUOVNEF
8	LogP_DTXSID_neighbor_1	DTXSID7020182
9	LogP_DTXSID_neighbor_2	DTXSID5087493C
0	LogP_DTXSID_neighbor_3	DTXSID50100531
1	LogP_DTXSID_neighbor_4	DTXSID0037231
2	LogP_DTXSID_neighbor_5	DTXSID4020412E
3	LogP_Exp_neighbor_1	3.32
4	LogP_Exp_neighbor_2	3.25
5	LogP_Exp_neighbor_3	2.11
5	LogP_Exp_neighbor_4	4.16
7	LogP_Exp_neighbor_5	2.69
8	LogP_pred_neighbor_1	3.18
9	LogP_pred_neighbor_2	3.66
0	LogP_pred_neighbor_3	3.5
1	LogP_pred_neighbor_4	3.05
2	LogP_pred_neighbor_5	3.44

National Institute of Environmental Health Sciences
Division of Translational Toxicology

QSAR-ready standardization tool





Command line

Graphical user interface

National Institute of Environmental Health Sciences
Division of Translational Toxicology

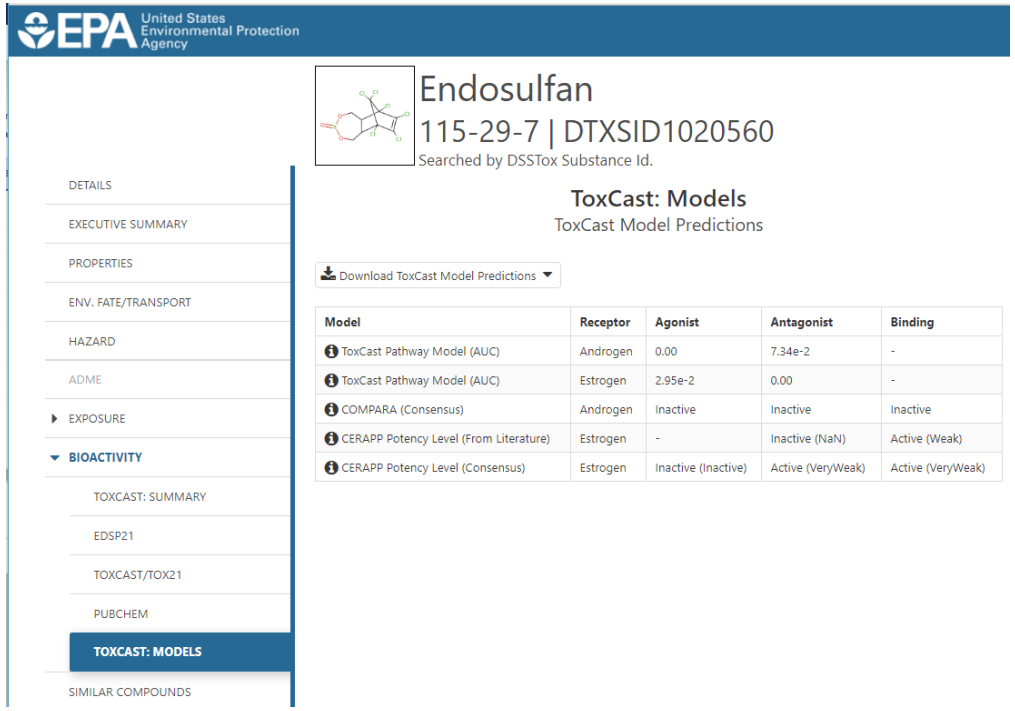
Predictions on NICEATM's ICE

https://ice.ntp.niehs.nih.gov/

National Institute of Environmental Health Sciences
Division of Translational Toxicology

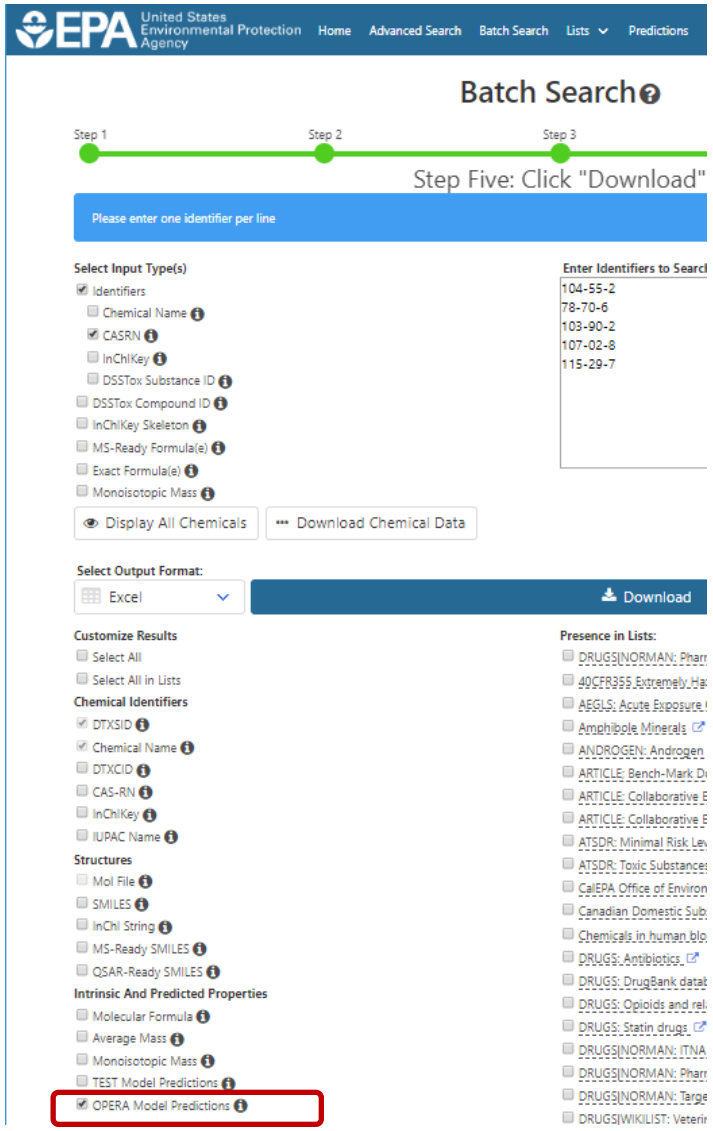
Predictions on the EPA Dashboard

Single chemical



https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID1020560#bioactivity-toxcast-models

Batch search



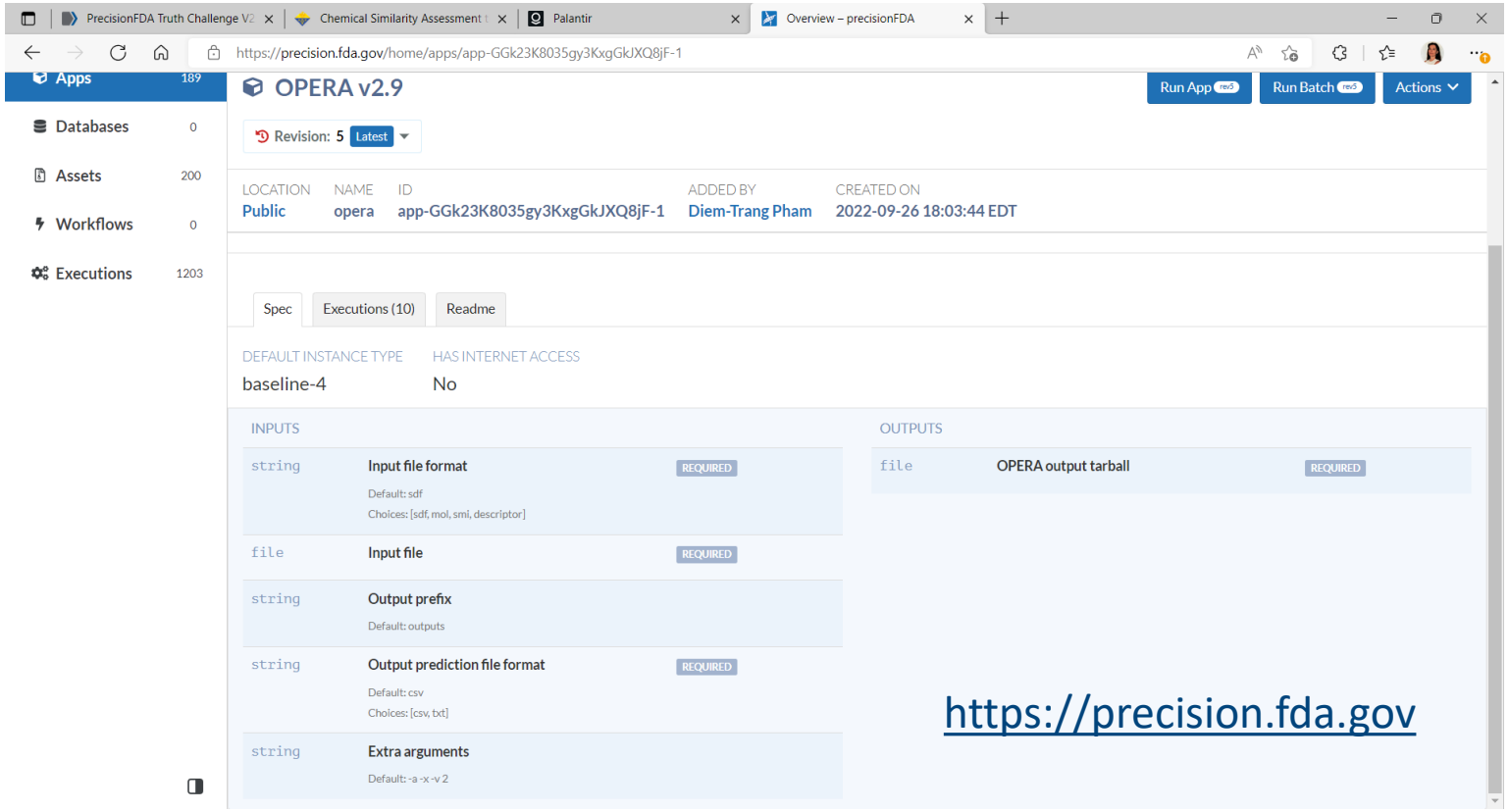
https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Predictions on FDA's precision platform

precisionFDA

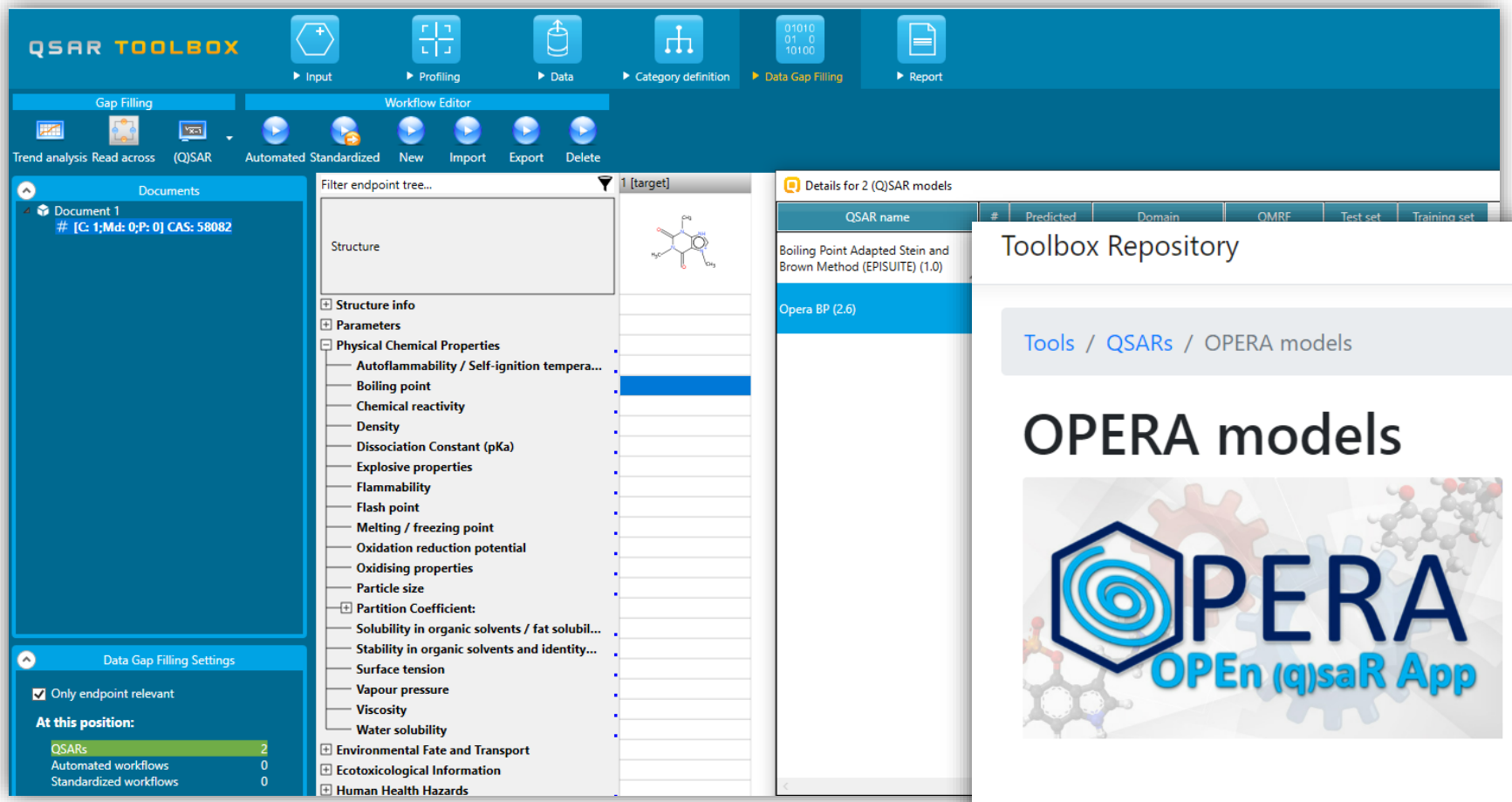
A secure, collaborative, high-performance computing platform that builds a community of experts around the analysis of biological datasets in order to advance precision medicine.

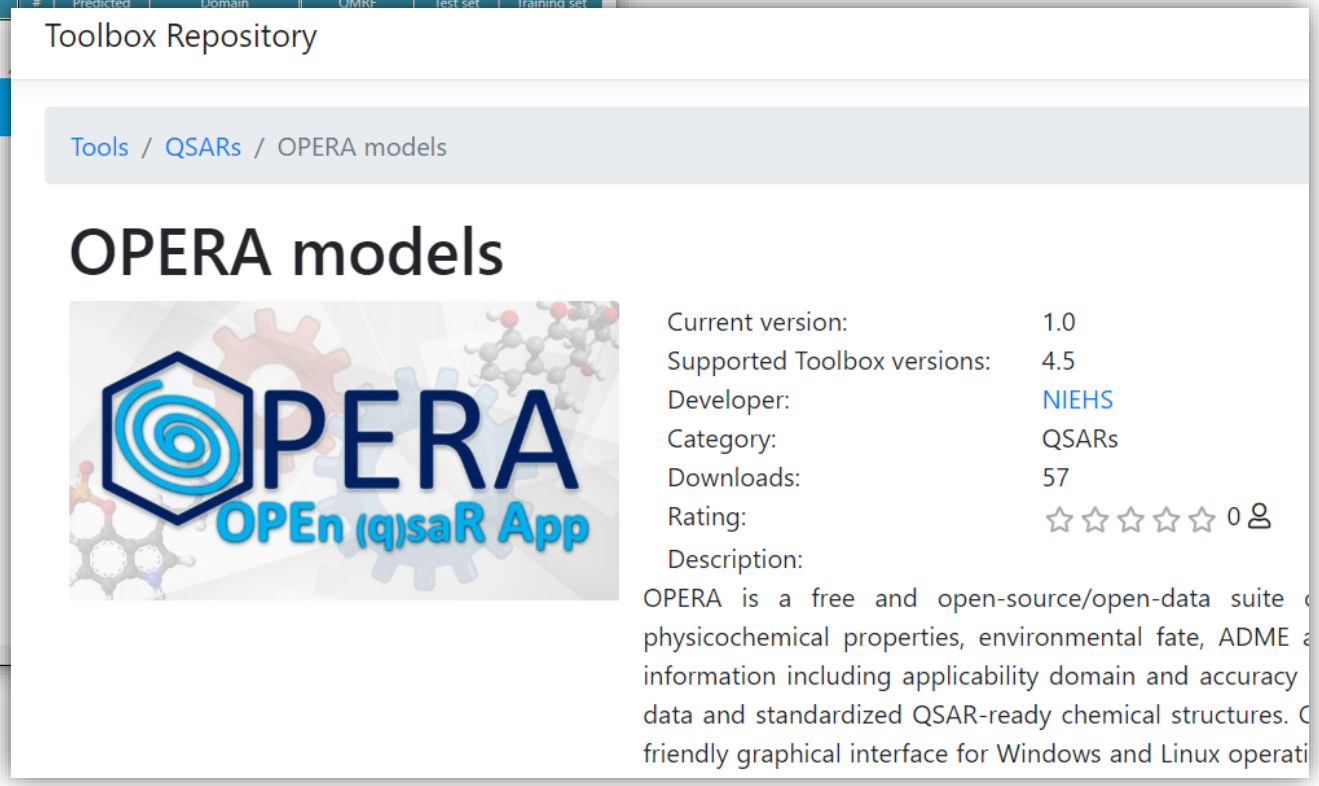


https://precision.fda.gov

National Institute of Environmental Health Sciences
Division of Translational Toxicology

OPERA in the OECD Toolbox





https://repository.qsartoolbox.org/

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Acknowledgments

The NICEATM Group

Subscribe to NICEATM News email list
https://ntp.niehs.nih.gov/go/niceatm

US EPA
All international collaborators

https://github.com/NIEHS/OPERA

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Thank you for your attention!