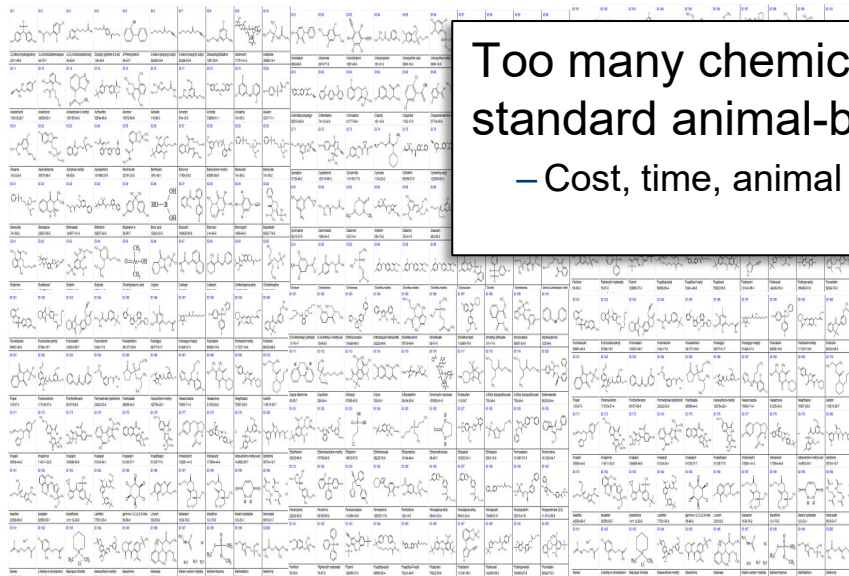


OPERA: Open-Source QSAR Models for Regulatory Support

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.



Too many chemicals to test with
standard animal-based methods
– Cost, time, animal welfare

Alternative

(Q)SAR
=
(Quantitative) Structure-Activity Relationship

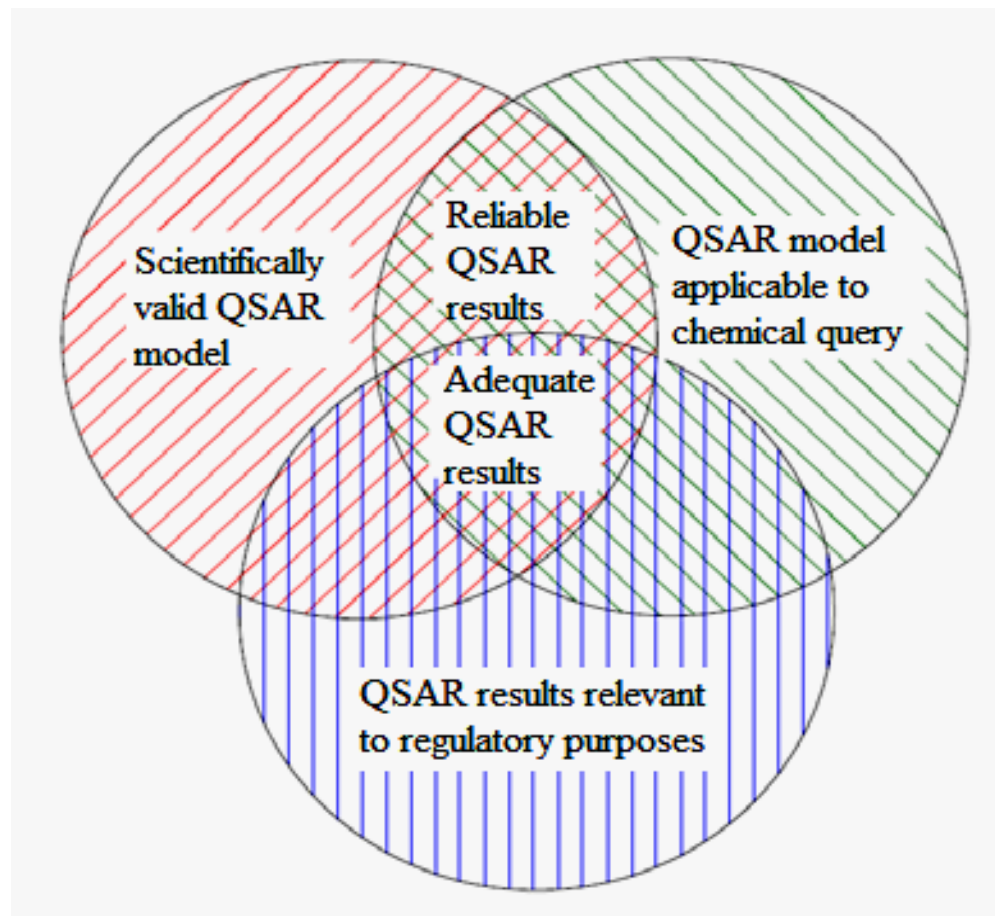


IN SILICO

- 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.

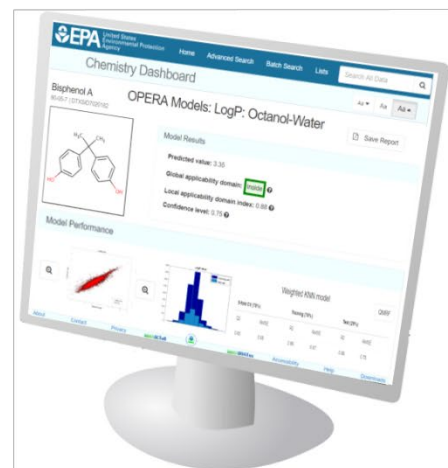
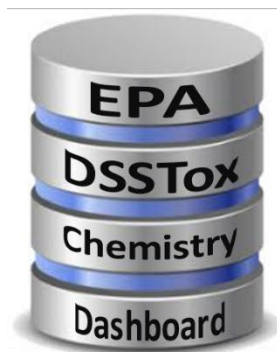
The 5 OECD Principles

- 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



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**

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

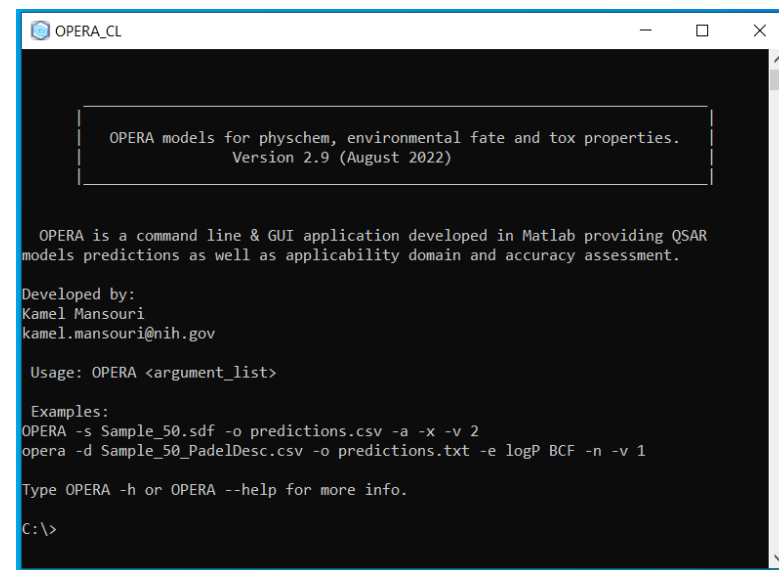
Input options:

- 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>



```
OPERA_CL
-----
OPERA models for physchem, environmental fate and tox properties.
Version 2.9 (August 2022)

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

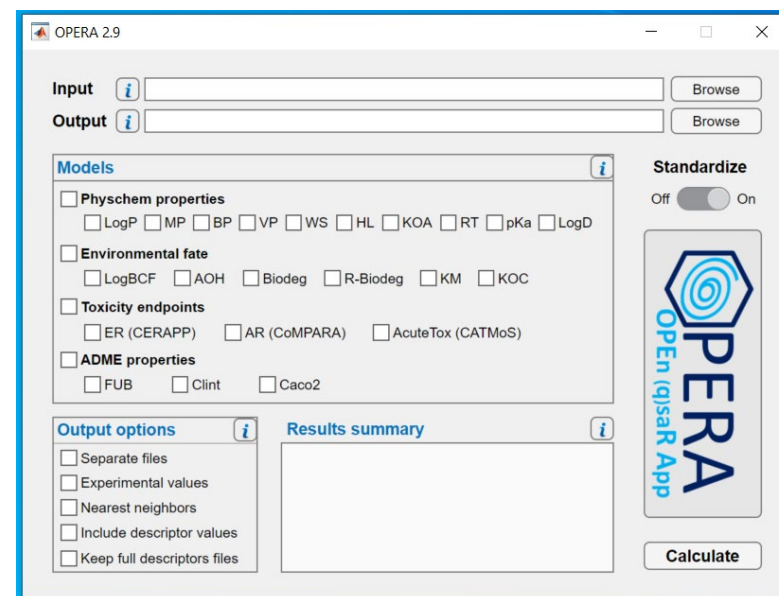
Developed by:
Kamel Mansouri
kamel.mansouri@nih.gov

Usage: OPERA <argument_list>


Examples:
OPERA -s Sample_50.sdf -o predictions.csv -a -x -v 2
opera -d Sample_50_PadelDesc.csv -o predictions.txt -e logP BCF -n -v 1


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


C:\>
```



OPERA 2.9

Input 

Output 

Models 

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 Caco2

Standardize

Off On

 OPERA
Open (q)SAR App

Output options 


Separate files

Experimental values

Nearest neighbors

Include descriptor values

Keep full descriptors files

Results summary 

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 $\log_{10} p$ (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 Pamis*

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

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

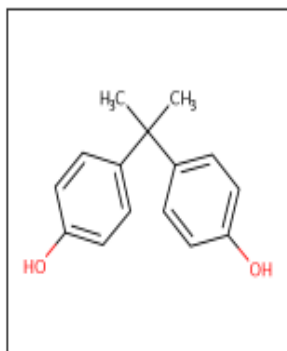


Acute Toxicity Workgroup: alternative methods

ICCVAM: Interagency Coordinating Committee on the Validation of Alternative Methods

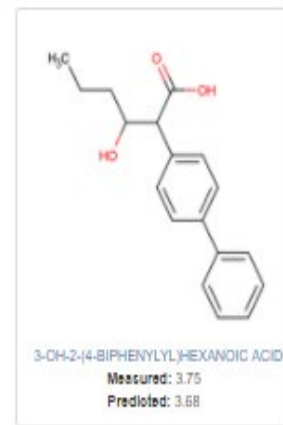
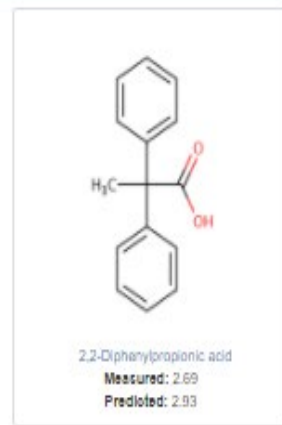
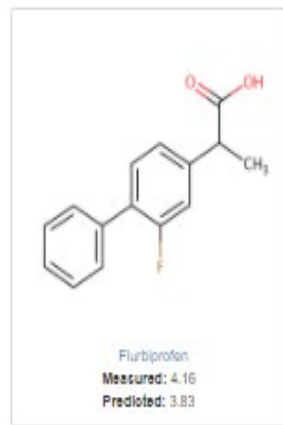
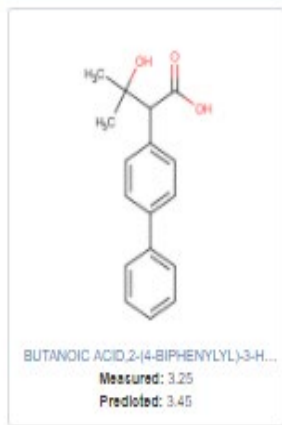
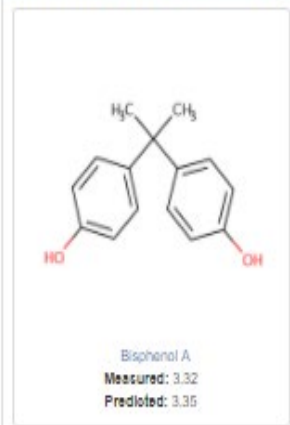
OPERA output

Bisphenol A^{MS}
 80-05-7 | DTXSID7020182



- 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

Nearest Neighbors from the Training Set



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
LogP_CAS_neighbor_3	85-27-8
LogP_CAS_neighbor_4	5104-49-4
LogP_CAS_neighbor_5	5558-66-7
LogP_InChiKey_neighbor_1	IISBACLAFKSPIT-I
LogP_InChiKey_neighbor_2	ZNPAJPOVURPFI
LogP_InChiKey_neighbor_3	PQSNIMHIHYFE
LogP_InChiKey_neighbor_4	SYTBZMRGLBWN
LogP_InChiKey_neighbor_5	ODELFXUOVNEF
LogP_DTXSID_neighbor_1	DTXSID7020182
LogP_DTXSID_neighbor_2	DTXSID5087493C
LogP_DTXSID_neighbor_3	DTXSID50100531
LogP_DTXSID_neighbor_4	DTXSID0037231
LogP_DTXSID_neighbor_5	DTXSID40204126
LogP_Exp_neighbor_1	3.32
LogP_Exp_neighbor_2	3.25
LogP_Exp_neighbor_3	2.11
LogP_Exp_neighbor_4	4.16
LogP_Exp_neighbor_5	2.69
LogP_pred_neighbor_1	3.18
LogP_pred_neighbor_2	3.66
LogP_pred_neighbor_3	3.5
LogP_pred_neighbor_4	3.05
LogP_pred_neighbor_5	3.44

QSAR-ready standardization tool

```
OPERA_CL
C:\Users\kmansouri\Downloads>opera -s PestActivesEPA.smi --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 Browse

Output Browse

Models

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

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

Results summary

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.

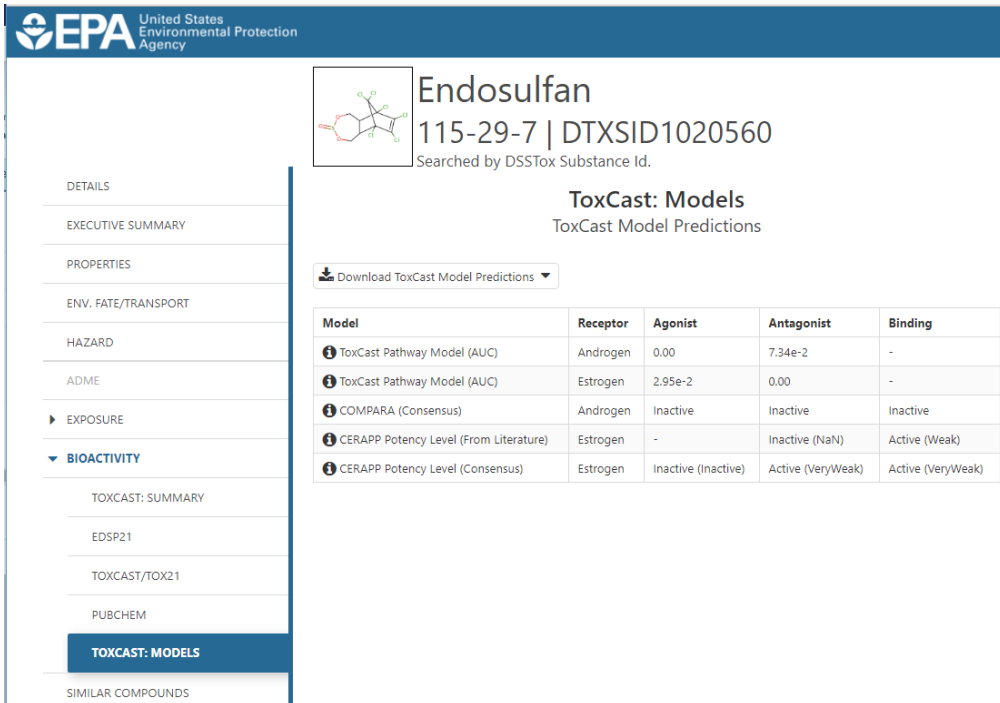
Standardize
Off On

OPERA
Open (q) SAR App

Calculate

Graphical user interface

Single chemical



EPA United States Environmental Protection Agency

Endosulfan
 115-29-7 | DTXSID1020560
 Searched by DSSTox Substance Id.

ToxCast: Models
ToxCast Model Predictions

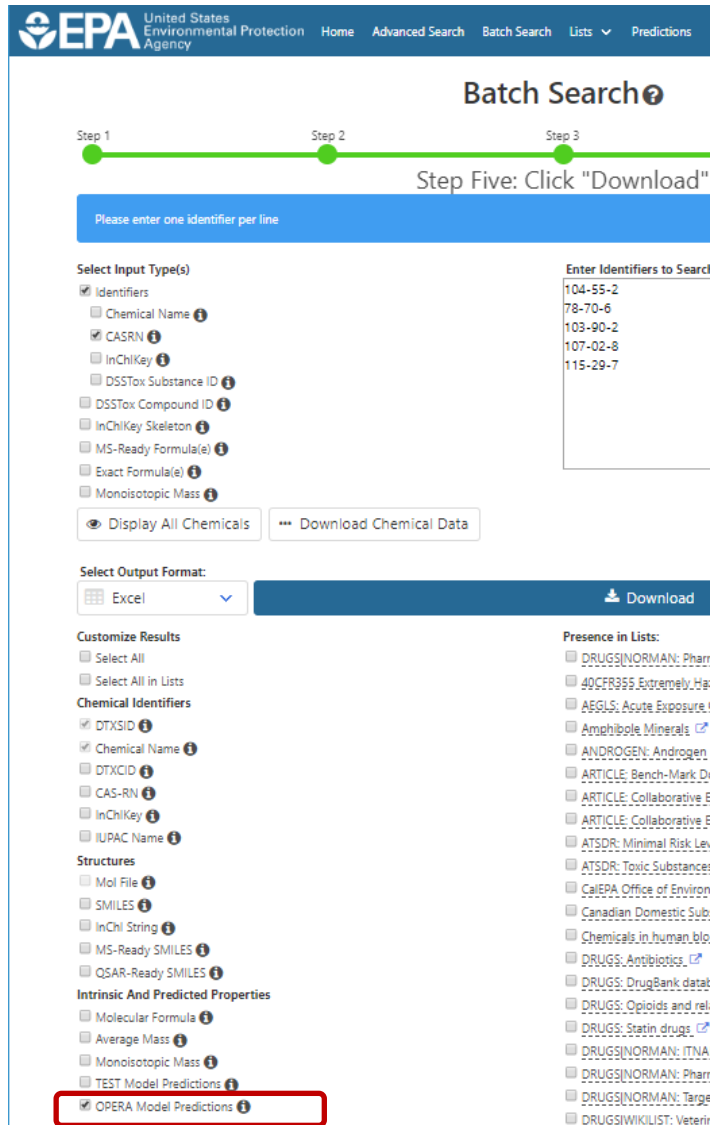
Download ToxCast Model Predictions

Model	Receptor	Agonist	Antagonist	Binding
ToxCast Pathway Model (AUC)	Androgen	0.00	7.34e-2	-
ToxCast Pathway Model (AUC)	Estrogen	2.95e-2	0.00	-
COMPARA (Consensus)	Androgen	Inactive	inactive	Inactive
CERAPP Potency Level (From Literature)	Estrogen	-	inactive (NaN)	Active (Weak)
CERAPP Potency Level (Consensus)	Estrogen	Inactive (Inactive)	Active (VeryWeak)	Active (VeryWeak)

DETAILS
 EXECUTIVE SUMMARY
 PROPERTIES
 ENV. FATE/TRANSPORT
 HAZARD
 ADME
 EXPOSURE
BIOACTIVITY
 TOXCAST: SUMMARY
 EDSP21
 TOXCAST/TOX21
 PUBCHEM
TOXCAST: MODELS
 SIMILAR COMPOUNDS

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

Batch search



EPA United States Environmental Protection Agency

Batch Search

Step 1 Step 2 Step 3
 Step Five: Click "Download"

Please enter one identifier per line

Select Input Type(s)
 Identifiers
 Chemical Name
 CASRN
 InChIKey
 DSSTox Substance ID
 DSSTox Compound ID
 InChIKey Skeleton
 MS-Ready Formula(e)
 Exact Formula(e)
 Monoisotopic Mass

Enter Identifiers to Search
 104-55-2
 78-70-6
 103-90-2
 107-02-8
 115-29-7

Select Output Format:
 Excel

Customize Results
 Select All
 Select All in Lists

Chemical Identifiers
 DTXSID
 Chemical Name
 DTXCID
 CAS-RN
 InChIKey
 IUPAC Name

Structures
 Mol File
 SMILES
 InChI String
 MS-Ready SMILES
 QSAR-Ready SMILES

Intrinsic And Predicted Properties
 Molecular Formula
 Average Mass
 Monoisotopic Mass
 TEST Model Predictions
 OPERA Model Predictions

Presence in Lists:
 DRUGSINORMAN: Pharr
 40CFR995: Extremely Ha
 AEGUS: Acute Exposure I
 Amphibole Minerals
 ANDROGEN: Androgen
 ARTICLE: Bench-Mark Di
 ARTICLE: Collaborative E
 ARTICLE: Collaborative E
 ATSDR: Minimal Risk Lev
 ATSDR: Toxic Substance
 CalEPA Office of Environ
 Canadian Domestic Sub
 Chemicals in human blo
 DRUGS: Antibiotics
 DRUGS: DrugBank data
 DRUGS: Opioids and rel
 DRUGS: Statin drugs
 DRUGSINORMAN: ITNA
 DRUGSINORMAN: Pharr
 DRUGSINORMAN: Targe
 DRUGSIWIKILIST: Veterir

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

Predictions on FDA's precision platform



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.

The screenshot shows a web browser window with the URL <https://precision.fda.gov/home/apps/app-GGk23K8035gy3KxgGkJXQ8jF-1>. The page displays the configuration for the OPERA v2.9 application. On the left, a sidebar menu shows 'Apps' (189), 'Databases' (0), 'Assets' (200), 'Workflows' (0), and 'Executions' (1203). The main content area shows the application name 'OPERA v2.9' with a 'Revision: 5 Latest' dropdown. Below this is a table with columns for LOCATION, NAME, ID, ADDED BY, and CREATED ON. The table contains one entry: Public, opera, app-GGk23K8035gy3KxgGkJXQ8jF-1, Diem-Trang Pham, 2022-09-26 18:03:44 EDT. There are 'Run App' and 'Run Batch' buttons. Below the table are tabs for 'Spec', 'Executions (10)', and 'Readme'. The 'Spec' tab is active, showing 'DEFAULT INSTANCE TYPE' as 'baseline-4' and 'HAS INTERNET ACCESS' as 'No'. The 'INPUTS' section includes: 'Input file format' (string, REQUIRED, Default: sdf, Choices: [sdf, mol, smi, descriptor]), 'Input file' (file, REQUIRED), 'Output prefix' (string, Default: outputs), 'Output prediction file format' (string, REQUIRED, Default: csv, Choices: [csv, txt]), and 'Extra arguments' (string, Default: -a -x -v 2). The 'OUTPUTS' section includes: 'OPERA output tarball' (file, REQUIRED). A URL <https://precision.fda.gov> is displayed in the bottom right corner of the screenshot.


The screenshot shows the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. Below this is a 'Workflow Editor' section with buttons for 'Automated', 'Standardized', 'New', 'Import', 'Export', and 'Delete'. The left sidebar shows a 'Documents' panel with 'Document 1' containing chemical structure information: '# [C: 1;Md: 0;P: 0] CAS: 58082'. The main workspace is divided into a 'Filter endpoint tree...' on the left and a 'Structure' view on the right showing a chemical structure. The 'Filter endpoint tree...' includes categories like 'Structure info', 'Parameters', 'Physical Chemical Properties' (with sub-items like Boiling point, Density, etc.), 'Partition Coefficient', and 'Environmental Fate and Transport'. The right panel shows 'Details for 2 (QSAR) models' with a table:

QSAR name	#	Predicted	Domain	OMRE	Test set	Training set
Boiling Point Adapted Stein and Brown Method (EPISUITE) (1.0)						
Opera BP (2.6)						

Toolbox Repository

Tools / QSARs / OPERA models

OPERA models



Current version: 1.0
Supported Toolbox versions: 4.5
Developer: NIEHS
Category: QSARs
Downloads: 57
Rating: ☆☆☆☆☆ 0

Description:
OPERA is a free and open-source/open-data suite of tools for predicting physicochemical properties, environmental fate, ADME and toxicological information including applicability domain and accuracy. It includes a user-friendly graphical interface for Windows and Linux operating systems.

<https://repository.qsartoolbox.org/>

The NICEATM Group



US EPA
All international collaborators



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



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

Thank you for your attention!



Question

OR



Comment