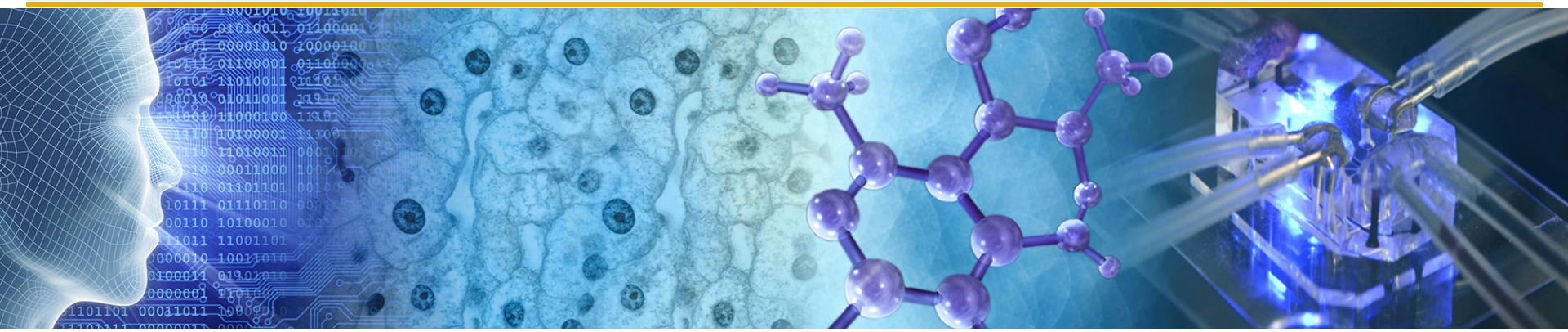
NTP
National Toxicology Program



# NICEATM Update

Nicole Kleinstreuer, PhD

Deputy Director, NICEATM

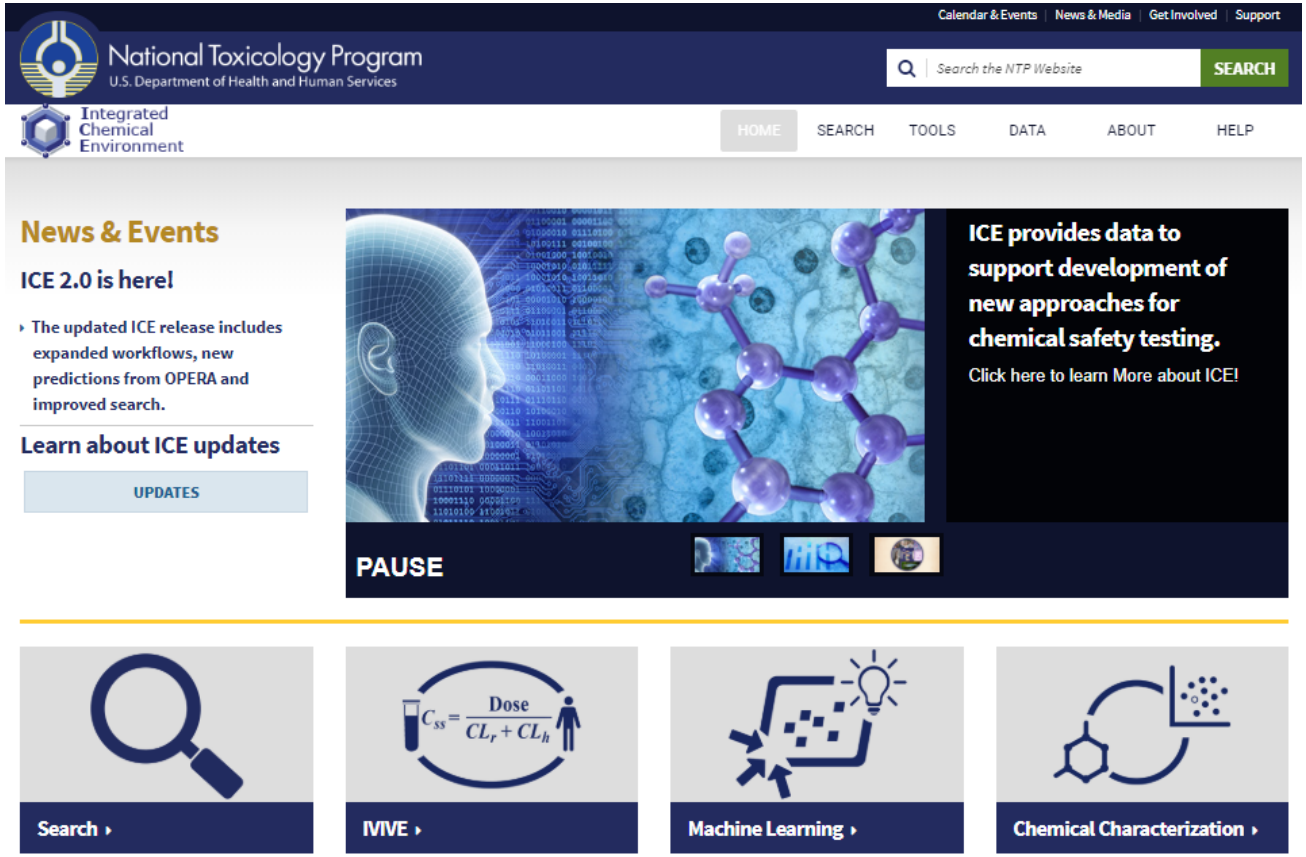
ICCVAM Public Forum
May 23, 2019

# Outline

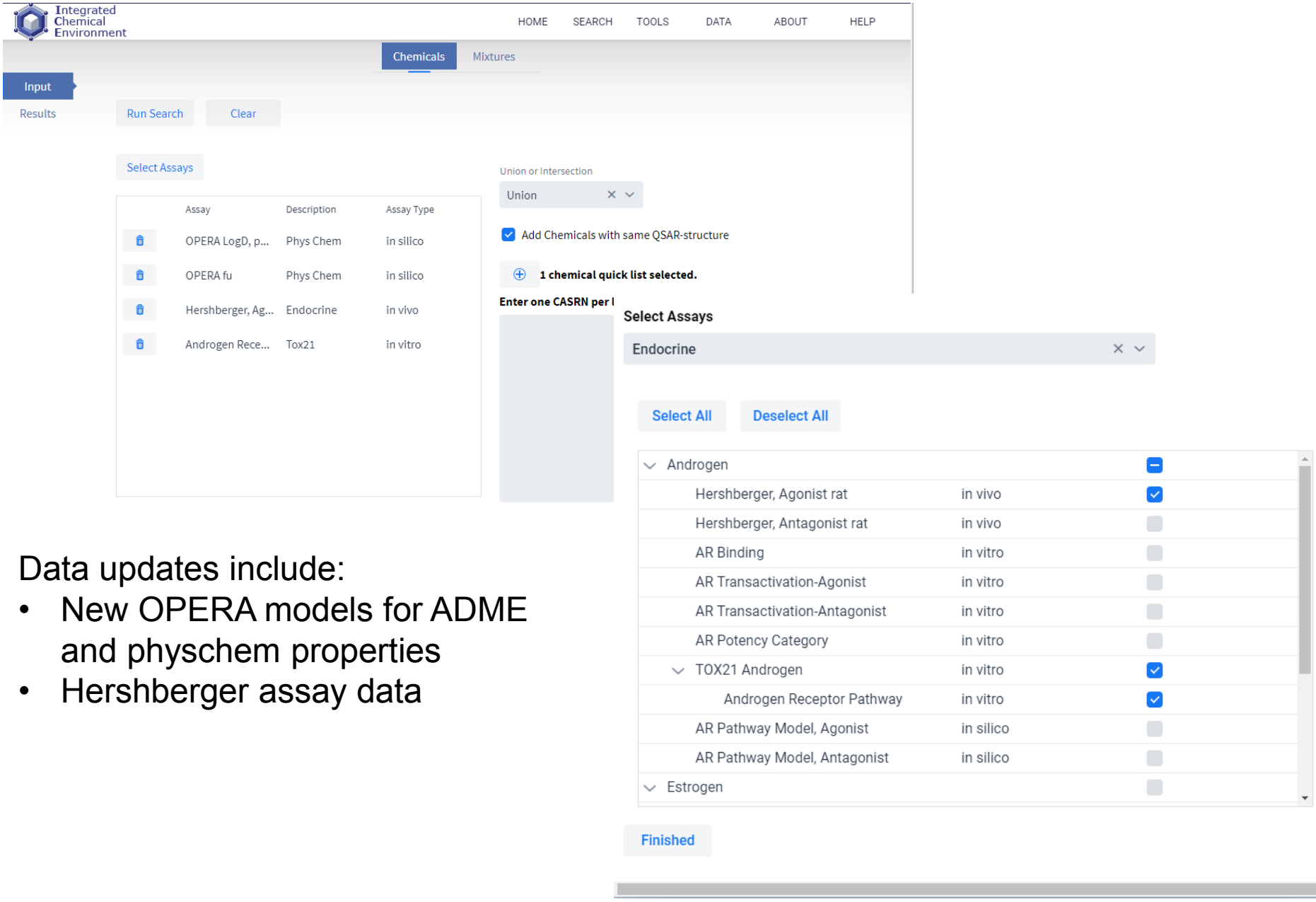
- ICE 2.0 and tools
  - CATMoS
  - IVIVE
  - Determining pKa from Chemical Structure
- OECD Skin Sensitization Project
- Automating Reference Data Identification for Developmental Toxicity
- Alternatives for Rabies Vaccine Testing
- Implementation of the Monocyte Activation Test for Medical Devices
- Alternatives for Botulinum Neurotoxin Testing

# Integrated Chemical Environment

- Updated Search
- Expanded IVIVE
- Data Updates



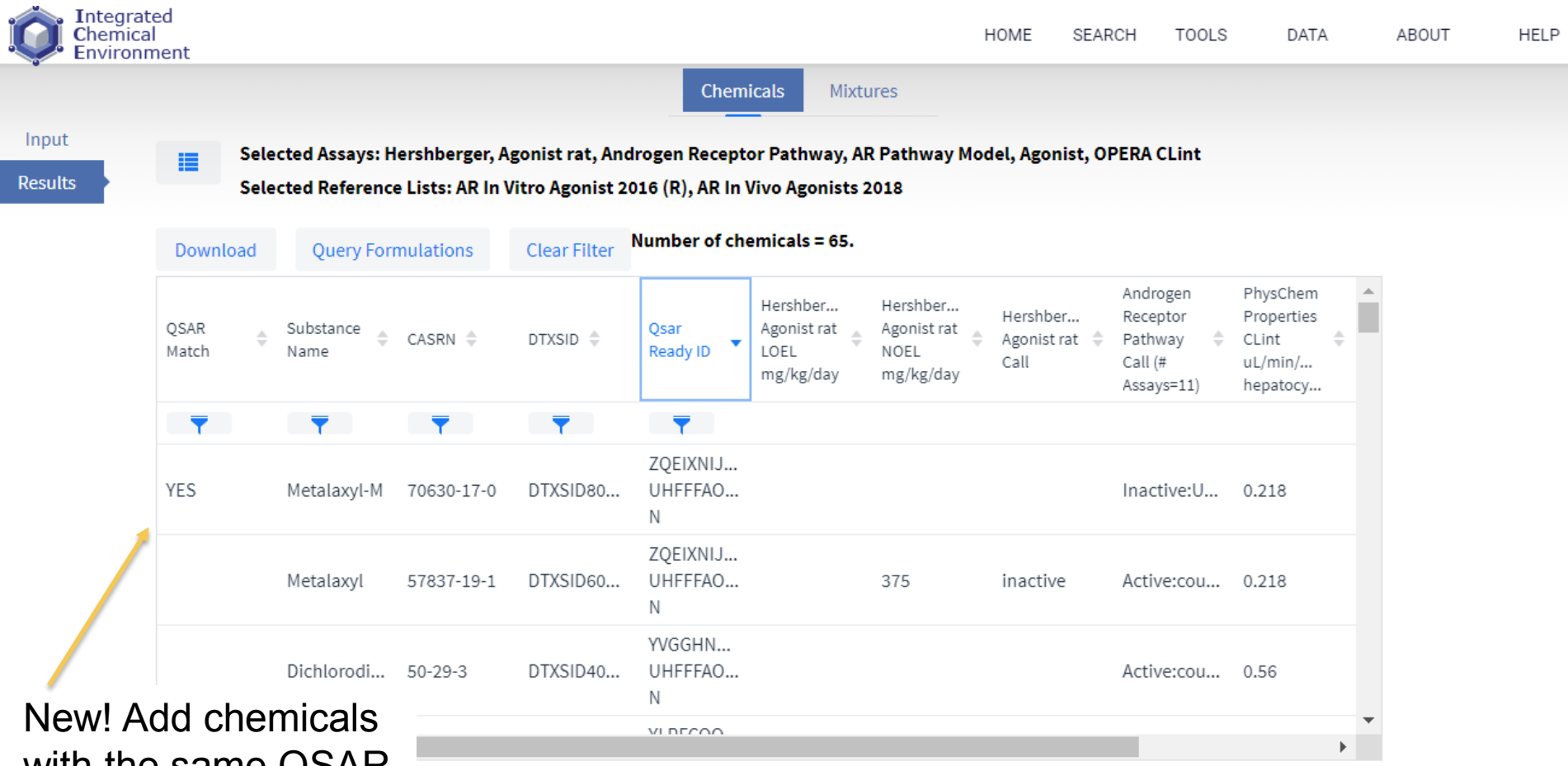
# Assay Selection




Data updates include:

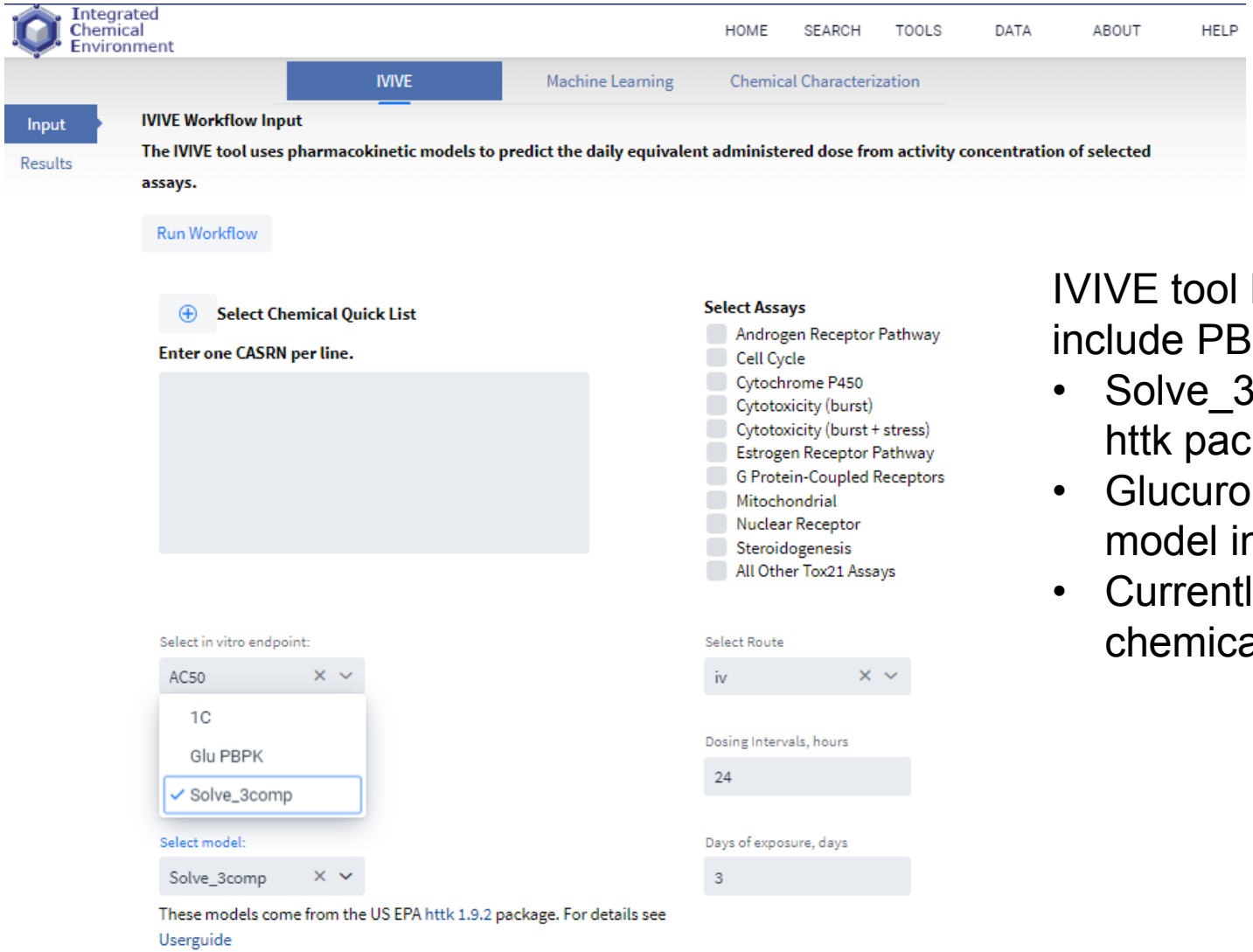
- New OPERA models for ADME and physchem properties
- Hershberger assay data

# Updated Search Table




New! Add chemicals with the same QSAR structure to query

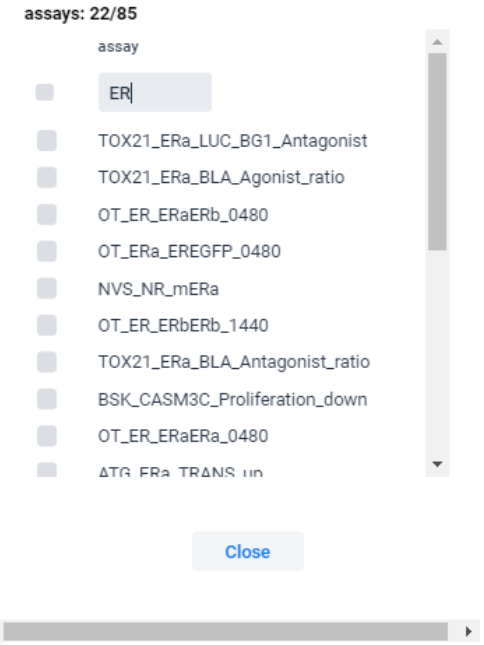
# Expanded IVIVE Tools




IVIVE tool has been updated to include PBPK models

- Solve_3comp from EPA's httk package
- Glucuronidation PBPK model in **BETA**
- Currently limited to 50 chemicals per query

# Overlay In Vivo Data on IVIVE Results


- Overlay in vivo data
- Update the plots with assay filtering
- Download data to explore locally

# Global Collaborative Projects

## CERAPP

Collaborative Estrogen Receptor Activity Prediction Project (2015/16)

## CoMPARA

**C**ollaborative **M**odeling **P**roject for **A**ndrogen **R**eceptor **A**ctivity **(2017/18)**

## CATMoS

**C**ollaborative **A**cute **T**oxicity **Mo**deling **S**uite **(2018/19)**

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

Endocrine Disruptor Screening Program (EDSP)

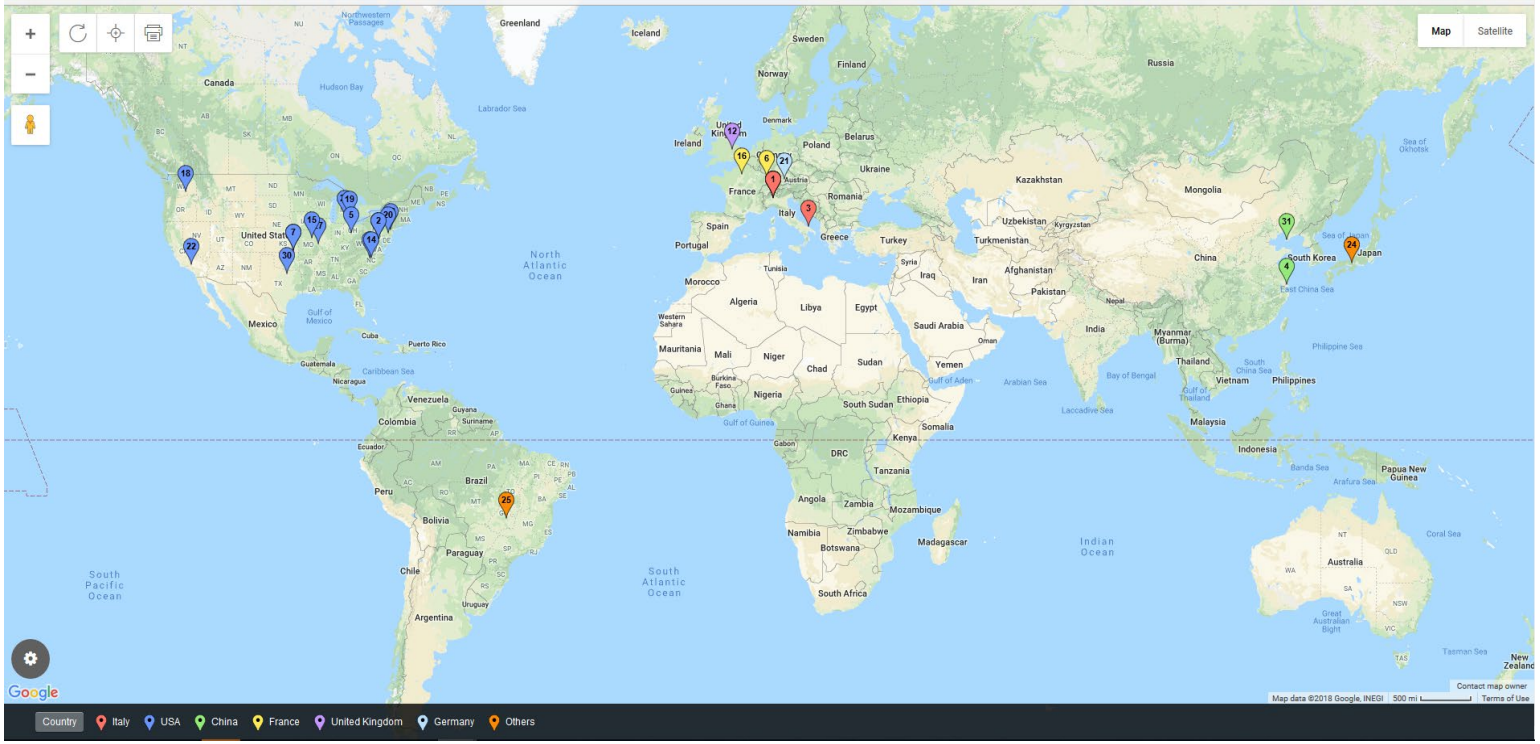

ICCVAM Acute Toxicity Workgroup

Over 100 international participants representing academia, industry, and government contributed.

# International Collaboration

## Consortium:

- **35 Participants/Groups** from around the globe representing academia, industry, and government contributed

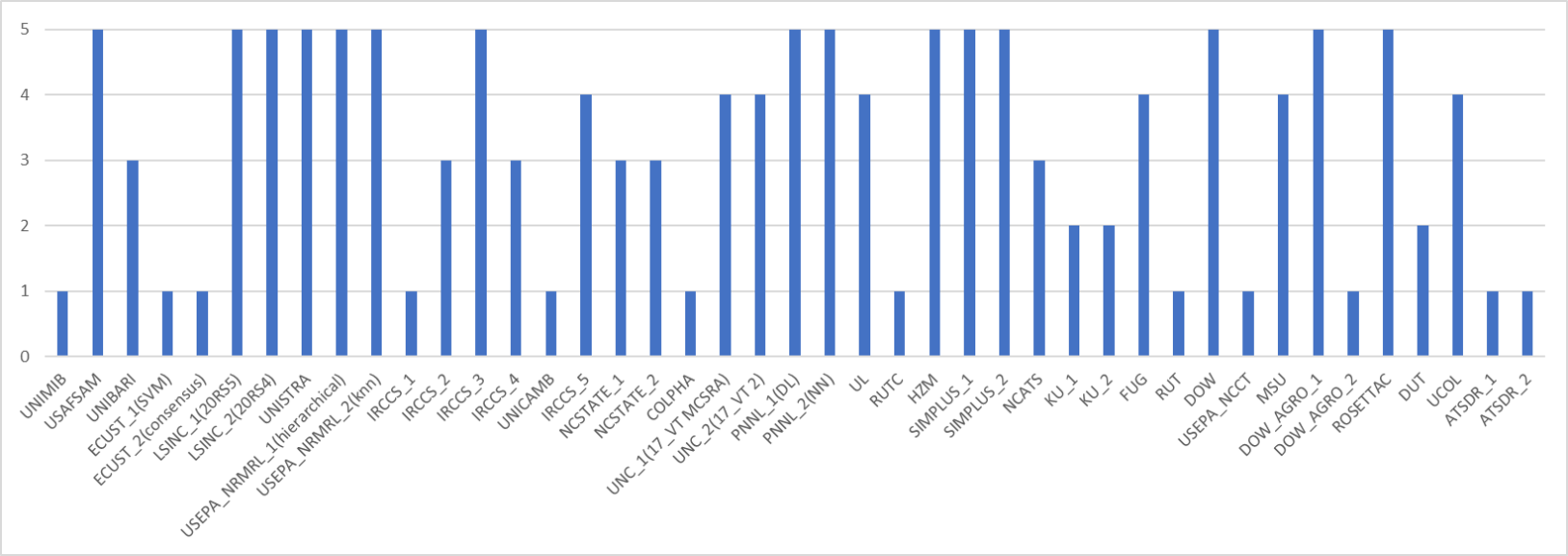



(https://batchgeo.com/map/d06c5d497ed8f76ecfee500c2b0e1dfa)

# CATMoS Submitted Models

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

**Total: 139 models**



# Evaluation procedure

## Qualitative evaluation:

- Documentation
- Defined endpoint
- Unambiguous algorithm
- Availability of code
- Applicability domain definition
- Availability of data used for modeling
- Mechanistic interpretation

## Quantitative evaluation:

- Goodness of fit: training statistics
- Evaluation set predictivity: statistics on the evaluation set
- Robustness: balance between (Goodness of fit) & (Test set predictivity)

$$S = 0.3 * (Goodness\ of\ fit) + 0.45 * (Test\ set\ predictivity) + 0.25 * (Robustness)$$

Categorical models (binary and multi-class):

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

$$Test\ set\ predictivity = 0.7 * (BA_{Tst}) + 0.3 * \left(1 - |\widetilde{Sn_{Tst} - Sp_{Tst}}|\right)$$

$$Robustness = 1 - |BA_{Tr} - BA_{Tst}|$$

Continuous models:

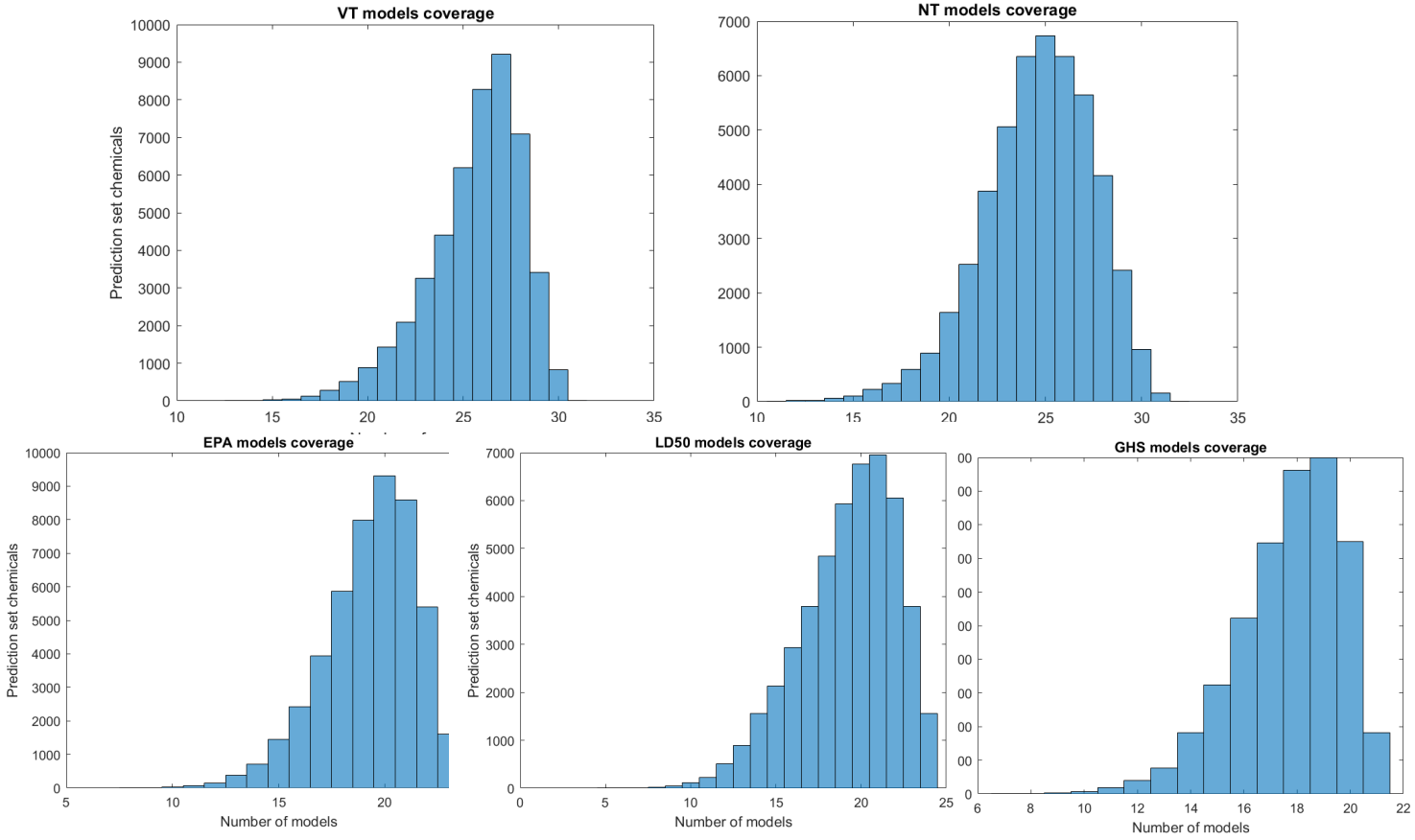
$$Goodness\ of\ fit = R^2_{Tr}$$

$$Test\ set\ predictivity = R^2_{Tst}$$

$$Robustness = 1 - |R^2_{Tr} - R^2_{Tst}|$$

# Coverage of the models

## Distribution of the number of models/chemical



# Generation of Consensus Predictions

- Models passing qualitative evaluation (requirement for transparency; description of approach was sufficient)
- Integrating only *in-domain* predictions across chemicals in the prediction set (48,137 chemicals) for each model, respectively
  - Categorical models: weighted majority rule
  - Continuous model: weighted average

# Performance Assessment

## CATMoS Consensus Model Statistics

| | Very Toxic | | Non-Toxic | | EPA | | GHS | |
|---|---|---|---|---|---|---|---|---|
| | Train | Eval | Train | Eval | Train | Eval | Train | Eval |
| Sensitivity | 0.87 | 0.67 | 0.93 | 0.70 | 0.73 | 0.50 | 0.63 | 0.45 |
| Specificity | 0.94 | 0.96 | 0.96 | 0.88 | 0.96 | 0.91 | 0.91 | 0.92 |
| Balanced Accuracy | 0.93 | 0.81 | 0.94 | 0.79 | 0.83 | 0.71 | 0.77 | 0.68 |
| ***In vivo* Balanced Accuracy** | 0.81 | | 0.89 | | 0.82 | | 0.79 | |

| | LD50 values | | LD50 values |
|---|---|---|---|
| | Train | Eval | ***In Vivo*** |
| R2 | 0.84 | 0.64 | 0.80 |
| RMSE | 0.32 | 0.51 | 0.42 |

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

# Consensus implementation

## Generalized CATMoS models: datasets

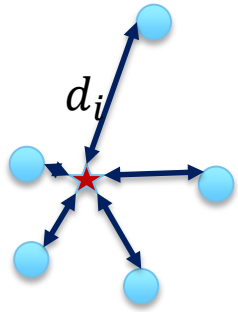
- LD50: 28954
- VT: 23767
- NT: 30971
- EPA: 25487
- GHS: 25720

- High concordance among models
- Proportional distribution of:
  - LD50 values
  - VT/NT classes
  - EPA/GHS categories

- Split into 75% training and 25% test set
- Calculate PaDEL & CDK2 descriptors
- Dimensionality reduction (missing values & low variance)
- Feature selection (most relevant descriptors for each endpoint)

# Consensus implementation

## Generalized CATMoS models: new chemical predictions




$d_1 \neq 0$

$w_i = f(d_i)$

$Pred_i = f(w_i, N_i)$

$d_1 = 0$

$Pred_i = N_i$

★ New chemical to be predicted

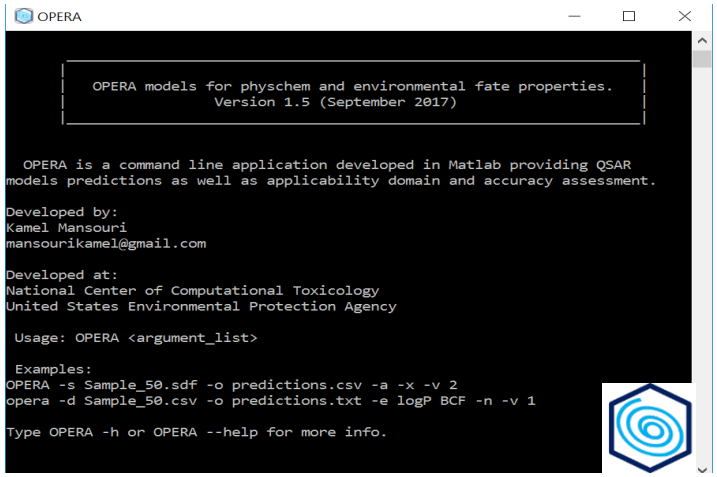
● Nearest neighbors ($N_i$)

$d_i$: Euclidean distance based on the selected descriptors for each endpoint

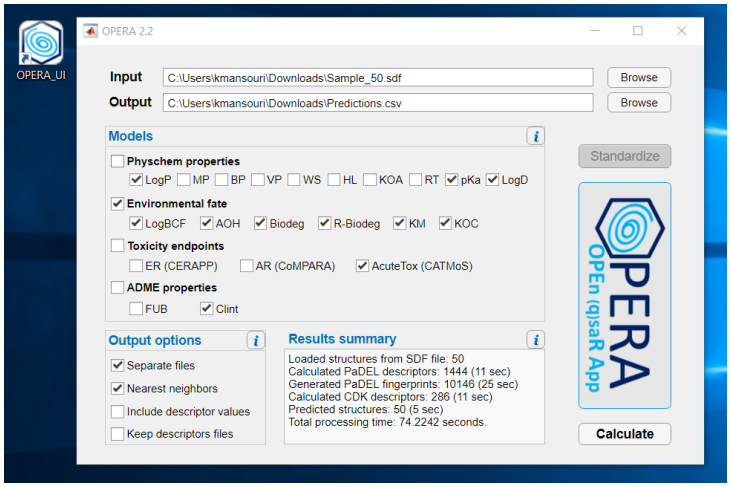
➡ Automated, weighted-endpoint dependent read-across: weighted kNN

# Running CATMoS Consensus models

## OPERA Standalone app




**Command line**




**Graphical user interface**

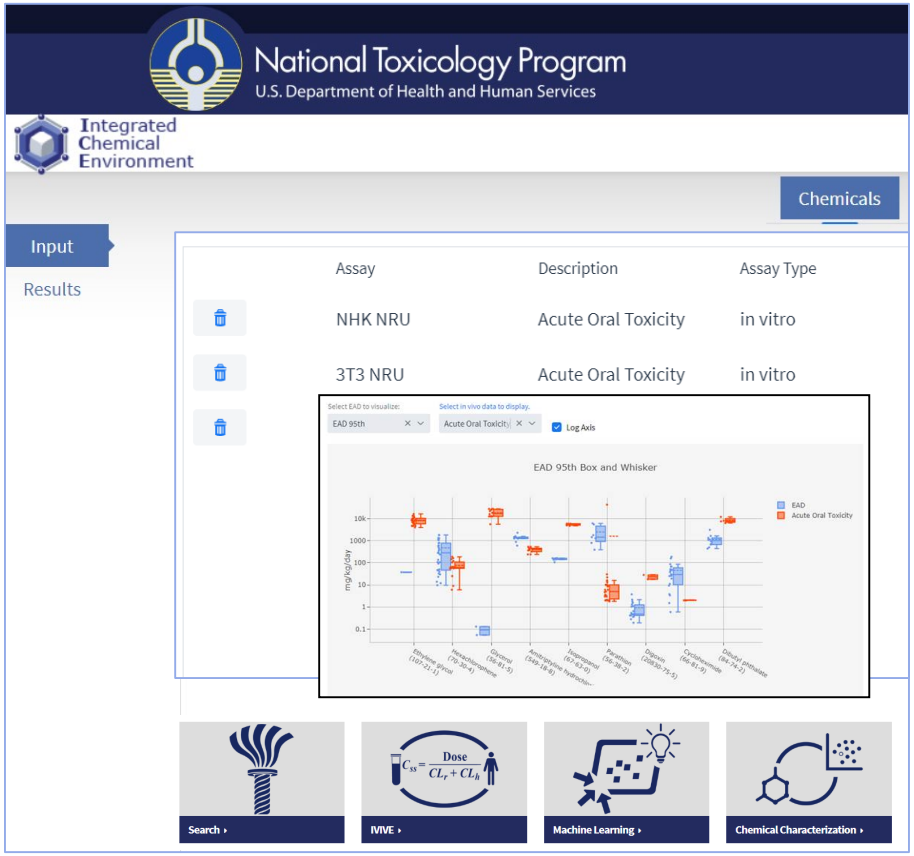
- Free, open-source & open-data
- Single chemical and batch mode
- Multiple platforms (Windows and Linux)
- Embeddable libraries (java, C, C++, Python)

**Mansouri et al. J Cheminform (2018). https://doi.org/10.1186/s13321-018-0263-1**

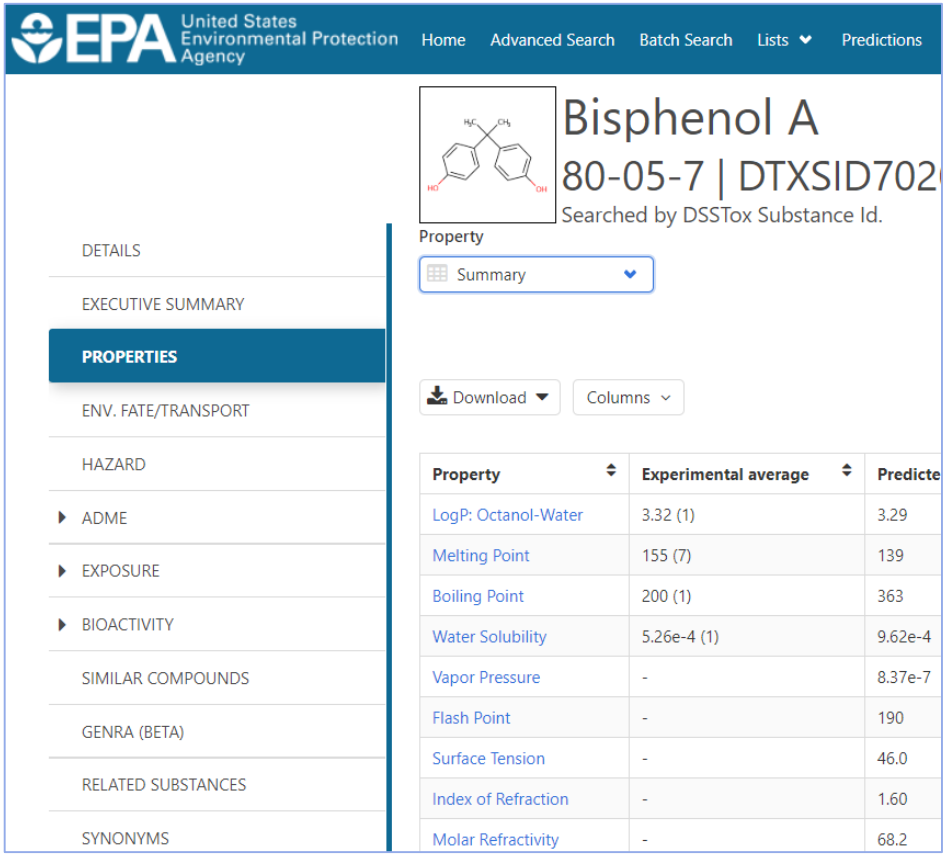
https://github.com/NIEHS/OPERA

# Available on ICE and EPA CompTox dashboard

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

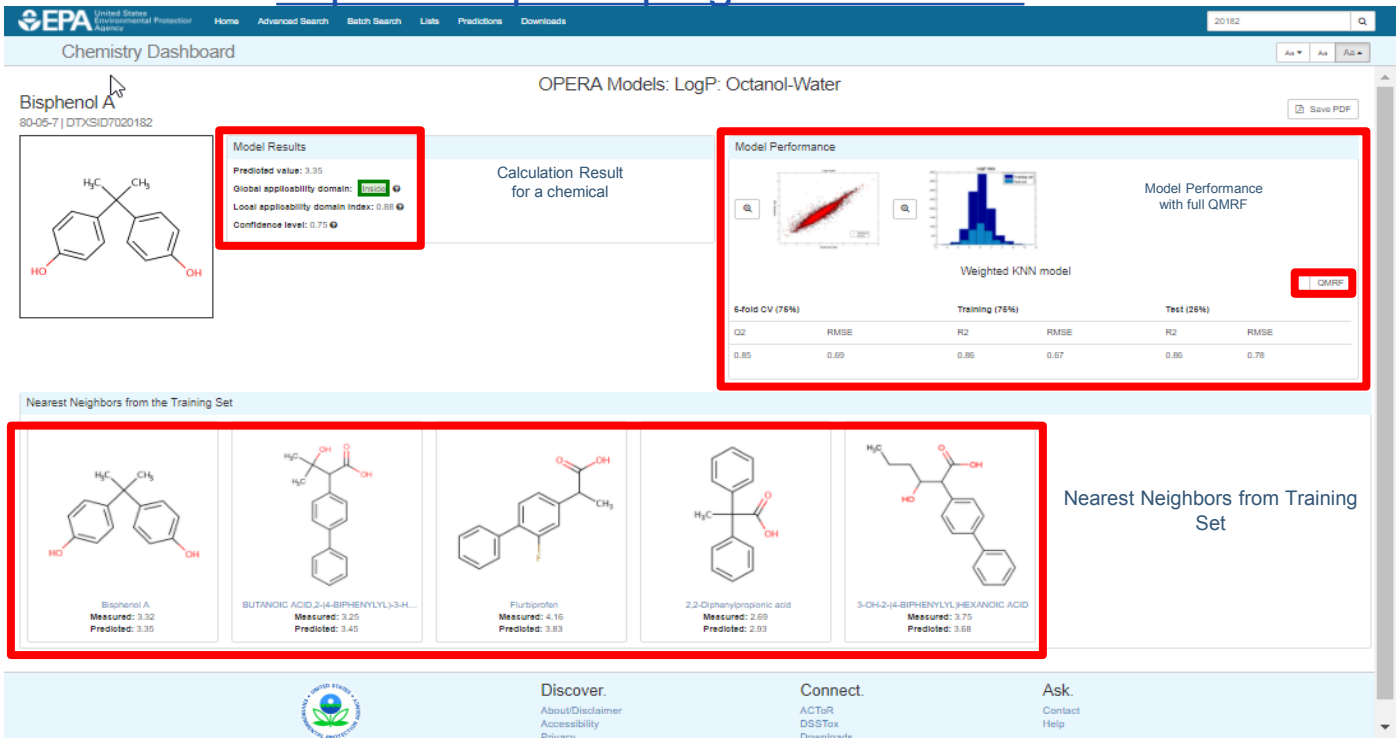



https://comptox.epa.gov/dashboard



# OPERA predictions on EPA's CompTox dashboard

https://comptox.epa.gov/dashboard




**Mansouri et al. OPERA models (https://doi.org/10.1186/s13321-018-0263-1)**
**Williams et al. CompTox Chemistry Dashboard (https://doi.org/10.1186/s13321-017-0247-6)**

## OPERA 1.5

### Physchem & Environmental fate:

| Model | Property |
|---|---|
| AOH | Atmospheric Hydroxylation Rate |
| BCF | Bioconcentration Factor |
| BioHL | Biodegradation Half-life |
| RB | Ready Biodegradability |
| BP | Boiling Point |
| HL | Henry's Law Constant |
| KM | Fish Biotransformation Half-life |
| KOA | Octanol/Air Partition Coefficient |
| LogP | Octanol-water Partition Coefficient |
| MP | Melting Point |
| KOC | Soil Adsorption Coefficient |
| VP | Vapor Pressure |
| WS | Water solubility |
| RT | HPLC retention time |

## New in OPERA2:

- Physchem properties:
  - General structural properties
  - pKa
  - Log D
- Toxicity endpoints
  - ER activity (CERAPP) https://ehp.niehs.nih.gov/15-10267/
  - AR activity (CoMPARA) https://doi.org/10.13140/RG.2.2.19612.80009
  - Acute toxicity (CATMoS) https://doi.org/10.1016/j.comtox.2018.08.002)
- ADME properties
  - Plasma fraction unbound (FuB)
  - Intrinsic clearance (Clint)

# Open Source QSAR Models For pKa

- The logarithmic dissociation constant, pKa, strongly influences a chemical's pharmacokinetic and biochemical properties:
  - Reflects the ionization state of a chemical,
  - Affects lipophilicity, solubility, protein binding, tissue:plasma partition coefficients and blood-brain barrier.

⇨ <u>pKa is important for ADMET properties, PBPK modeling and IVIVE</u>

**<u>Problem statement:</u>**
No currently available free and transparent predictors for heterogeneous chemical classes

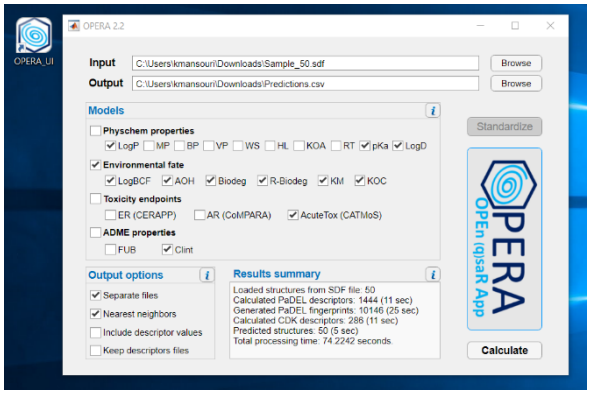
# pKa Modeling approach

Building free & open source models using free & open source tools:

- pKa values for 7912 chemicals collected from DataWarrior.
- Data curated and chemical structures were standardized for QSARs
- PaDEL software to calculate molecular descriptors and fingerprints.
- Several machine learning approaches were applied:
  - DNN: deep neural networks
  - SVM: support vector machine
  - XGB: extreme gradient boosting.
- Models were 5-fold cross-validated and evaluated against an external test set.
- The best models for each algorithm were compared to each other and to predictions from ACD/Labs and ChemAxon

# pKa project outcome

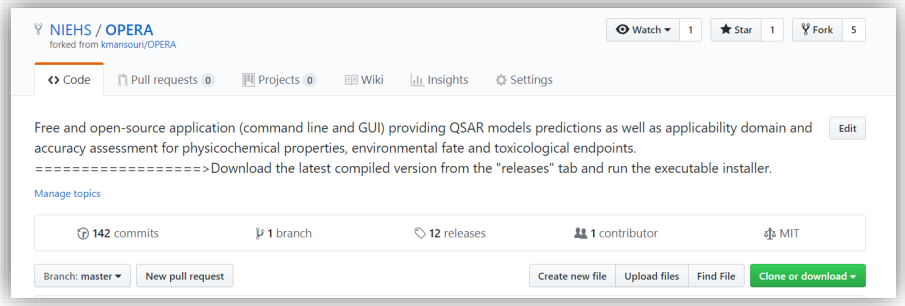
- Free and open source QSAR models for acidic and basic pKa
- Accuracy reaching an RMSE of 1.5 and an $R^2$ of 0.8
- Best models applied on DSSTox chemicals (~850k)
  - https://ice.ntp.niehs.nih.gov/
  - https://comptox.epa.gov/dashboard
- Manuscript submitted to Journal of Cheminformatics
- New chemicals can be predicted using OPERA (with applicability domain and accuracy estimates)




Code and executables freely available on

Github: https://github.com/NIEHS/OPERA

(Windows and Linux versions)



# Skin Sensitization Reference Databases

- In support of OECD Defined Approach Guideline project
- LLNA database
  - DASS Expert Group expressed concerns about reference classifications
  - Project leads are re-evaluating; eliminating tests without supporting concentration/SI data and those using modified protocols
- Human database
  - Being expanded and QC'd with help from German National Institute for Risk Assessment (BfR)

BfR
Bundesinstitut für Risikobewertung

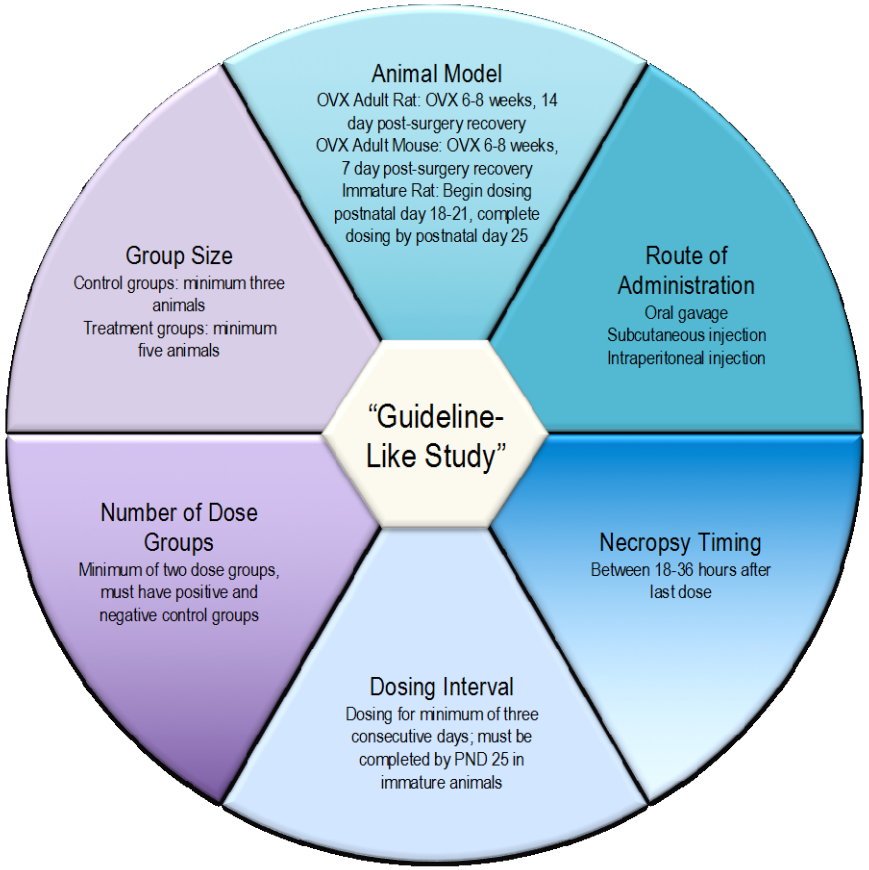
# Human Skin Sensitization Data Project

- Purpose: to provide reference human data to support the evaluation of alternative skin sensitization test methods
- Objectives
  1. Curate human predictive patch test database for DA performance assessment
  2. Analyze data to understand uncertainty and sources of variability
  3. Develop/apply a transparent, reproducible system for human skin sensitization potency categorization
     - Review/build on Basketter et al. 2014 & Api et al. 2017, which used data from predictive and diagnostic human patch tests

# Human Data Subgroup: Reviewers

- Anne Marie Api, Research Institute for Fragrance Materials
- Nicole Kleinstreuer, U.S. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
- Hon Sum Ko, U.S. Food and Drug Administration, Division of Dermatology and Dental Products
- Joanna Matheson, John Gordon, U.S. Consumer Product Safety Commission, Health Sciences Directorate
- Judy Strickland, Integrated Laboratory Systems, contractor supporting NICEATM
- Matthias Herzler, Herrmann-Josef Thierse, German Federal Institute for Risk Assessment (BfR), Chemical and Product Safety

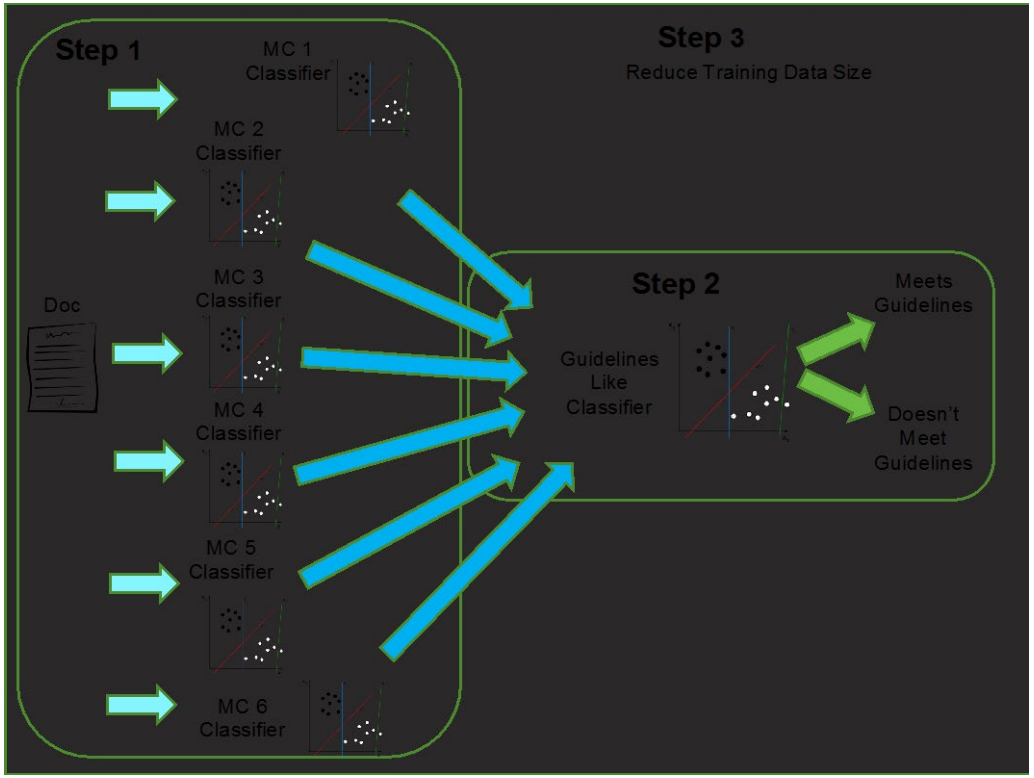
# Manually Identifying Reference Data




- Systematic literature search of publically available data (e.g. PubMed)
- Identify chemical activities measured in "guideline-like" uterotrophic studies
- Identify a subset of *in vivo* reference chemicals
  - Active chemicals verified in ≥2 independent studies
  - Inactive chemicals verified in ≥2 independent studies (with no positive results in any study)

*Kleinstreuer et al. EHP (2015)*

# Automating Reference Data Identification




- Project with Oak Ridge National Labs (ORNL) and FDA CFSAN to apply text-mining (NLP) approaches & ML to identify high-quality data
- Semi-automated retrieval and evaluation of published literature (trained on uterotrophic database)
- Apply to developmental toxicity studies

# DARTWG Activities

- Define literature search keywords
  - Use to search for developmental and reproductive toxicity studies
- Identify regulatory guidelines for conducting developmental and reproductive toxicity studies
- Extract study protocol details from guidelines
- Characterize study protocols to identify guideline components and minimum criteria
  - Group into tiered sets

    1) Required (appear in all TGs)

    2) Preferred

    3) Nice-to-have

A. Turley, US FDA postdoc

# Workshop on MAT for Medical Devices

- NICEATM and the PETA International Science Consortium (PISC) co-organized a workshop to discuss the use of the human cell-based monocyte activation test (MAT) as a standalone release test for medical devices.
  - September 18-19, 2018; NIH, Bethesda, MD, USA
- ~50 regulators, test developers, medical device manufacturers
  - discussed approaches to support use of the MAT for batch release testing of medical devices.
- Participants recommended studies needed to fill information gaps.
- Stakeholder subgroup will be convened to use the workshop conclusions as the basis for a proposal to the U.S. Food and Drug Administration's Medical Device Development Tools Program to consider the MAT as a non-animal alternative for pyrogen testing
- Finalizing workshop report for submission to ALTEX in the coming weeks

# Workshop on Rabies Vaccine Testing

- NICEATM and the International Alliance for Biological Standardization – North America (IABS-NA) co-organized workshop.
  - October 16-17, 2018; NIH, Bethesda, MD, USA
- ~60 scientists from government, academia and industry developed recommendations to advance alternative methods for human and veterinary rabies virus vaccine testing.
- Detailed the current state of the science of nonanimal alternatives to traditional animal-based rabies virus vaccine potency and safety tests.
- Breakout group discussions focused on the steps necessary for implementing alternatives for veterinary and human rabies virus vaccine potency testing.
- Identified actions and data needed for further progress
- Finalizing workshop report to summarize meeting outcomes and associated conclusions – will be submitted to the Biologicals for publication in the next few weeks.

# Non-animal approaches to detect botulinum neurotoxin

- Current assay to detect BoNT is a mouse bioassay
- A BoNT/E-specific ELISA-based kit has been developed by BioSentinel to detect BoNT/E in avian blood samples.
  - Current effort focused on establishing transferability to the National Wildlife Health Center (NWHC), part of the U.S. Geological Survey; also demonstrating long-term performance and suitability
- Additionally, a BoNT/C-specific ELISA-based kit has also been developed by BioSentinel.
  - Initial studies indicated that BoNT/C could not be detected in avian blood samples
  - Follow-up spike-recovery testing indicated that components in the blood were interfering with detection.
  - Current effort focused on additional development to reduce assay interference effects
- Collaborative effort with BioSentinal and NWHC

# Non-animal Affinity Reagents

- Challenge: Identifying alternatives to current animal-based methods for producing monoclonal and polyclonal antibodies (i.e., development of non-animal affinity reagents)
- EURL-ECVAM Scientific Advisory Committee (ESAC) reviewed the scientific validity of antibodies and non-antibody affinity reagents generated using animal-free technologies for use in research and diagnostics
- The ESAC WG has circulated their complete report for internal commenting and finalization.
  - Publication of the ESAC Opinion and Final Report scheduled for Spring 2019
- In general, there is a lack of any scientific evidence that these methods cannot be used in place of animal-based antibodies
- Based on the ESAC report, next steps for ICCVAM?
  - Facilitate implementation
  - Identify barriers and actionable solutions
  - Planned workshop for December 2019 (NIH – Porter Neuroscience Center)

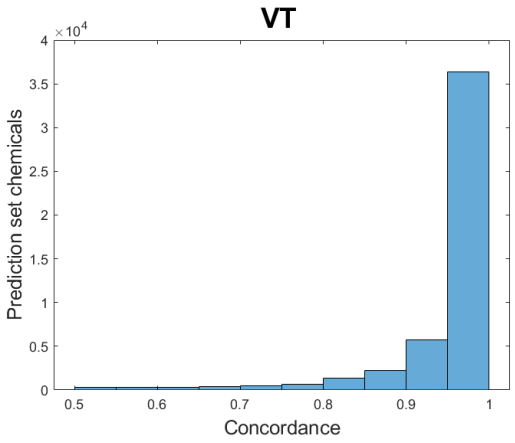
# Acknowledgments

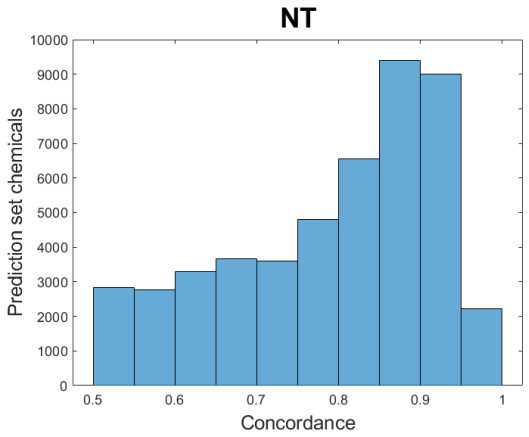

- ILS/NICEATM
- ICCVAM partners
- ICATM partners
- Modeling consortium participants

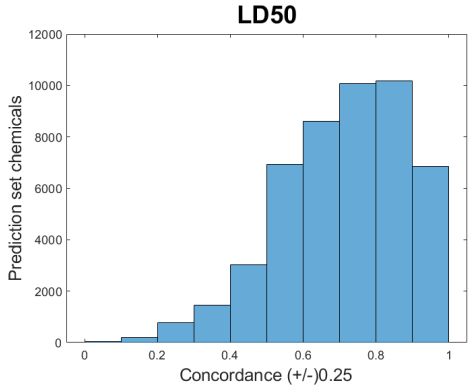

# Extra Slides

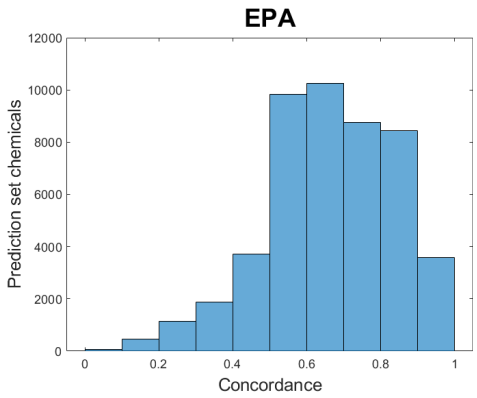
# Consensus concordance

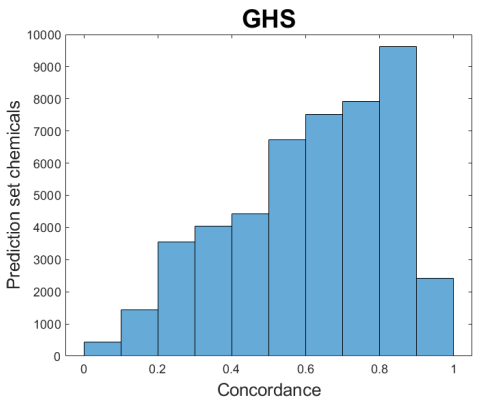
## Distributions of the concordance between models



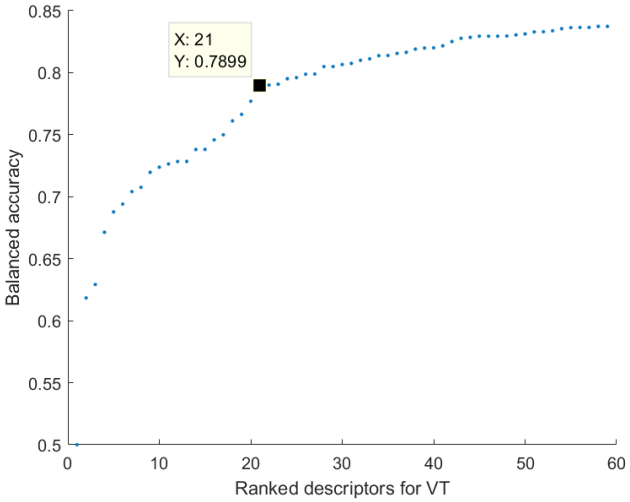



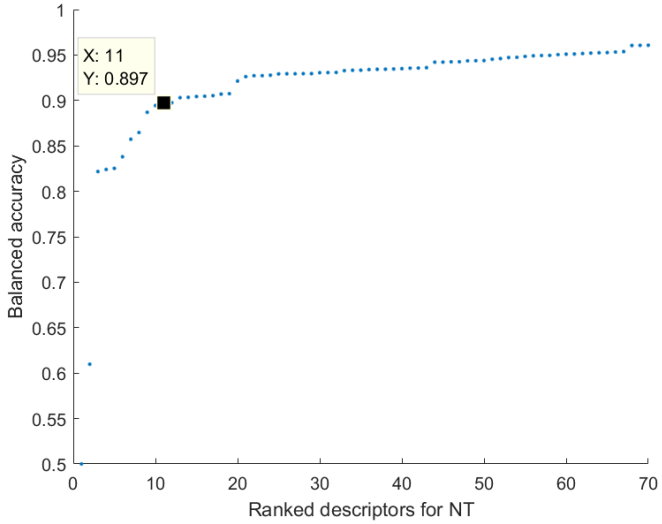



# Consensus implementation

## Generalized CATMoS models: statistics




| Endpoint | Descriptors | Training (5-f CV) | Test set |
|---|---|---|---|
| VT (BA) | 21 | 0.79 | 0.77 |
| NT (BA) | 11 | 0.90 | 0.89 |
| EPA (BA) | 15 | 0.79 | 0.81 |
| GHS (BA) | 15 | 0.78 | 0.79 |
| LD50 ($Q^2$,$R^2$) | 23 | 0.79 | 0.81 |