National Institute of Environmental Health Sciences
Division of the National Toxicology Program

OPERA Models for ADME-related Properties

K Mansouri[1], N Choksi[2], X Chang[2], D Allen[2], and N Kleinstreuer[1]
[1]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; [2]Inotiv, RTP, NC, USA

Abstract: 4654
Poster: P700

Introduction

- OPERA – the Open (Quantitative) Structure–activity/property Relationship App – is a free and open-source/open-data suite of quantitative structure–activity relationship (QSAR) models developed to support a range of research and regulatory purposes.
- In addition to physicochemical and environmental fate properties, OPERA offers a number of models predicting endpoints affecting absorption, distribution, metabolism, and excretion (ADME) that are of high importance to physiologically based kinetic (PBK) modeling and in vitro to in vivo extrapolation (IVIVE) studies.
- The OPERA ADME related endpoints include models for:
  - Physicochemical parameters: octanol/water partition (logP) and distribution (logD) coefficients, acidic dissociation (pKa).
  - Pharmacokinetic parameters: fraction unbound to plasma protein (F_u), intrinsic hepatic clearance (Cl_{int}), and the newly added Caco2 permeability (logPapp).
- OPERA predictions for ADME related properties are used as inputs for the IVIVE workflow on the National Toxicology Program (NTP) Integrated Chemical Environment.

OPERA Application

General approach:

- The five Organisation for Economic Co-operation and Development (OECD) principles for QSAR validation are employed during modeling.
- Only high-quality curated data are used to build the models.
- Chemical structures are processed using QSAR-ready standardization workflow.
- The QSAR-ready workflow is also implemented in the app for user input processing structures prior to prediction.
- Works with different input and output formats.
- OECD-compliant QSAR model reporting format (QMRF) reports are available.
- OPERA provides:
  - Applicability domain and prediction accuracy assessment.
  - Experimental values when available.
  - Information about the nearest neighbors.
  - Molecular descriptor values for transparency.

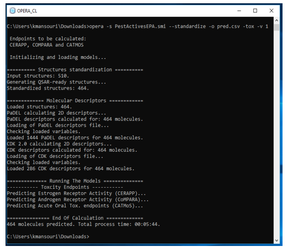
Availability:

Predictions:
- U.S. Environmental Protection Agency (EPA) CompTox Chemicals Dashboard: https://comptox.epa.gov/dashboard
- NTP Integrated Chemical Environment: https://ice.ntp.niehs.nih.gov/

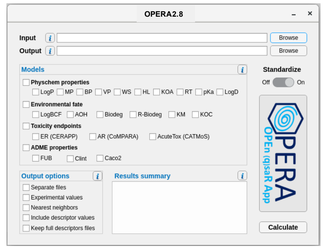
Standalone desktop application:
- Github: https://github.com/NIEHS/OPERA
- NTP KNIME server: knime.niehs.nih.gov/knime/

More info:
- https://ntp.niehs.nih.gov/go/opera

Interfaces (v2.8):

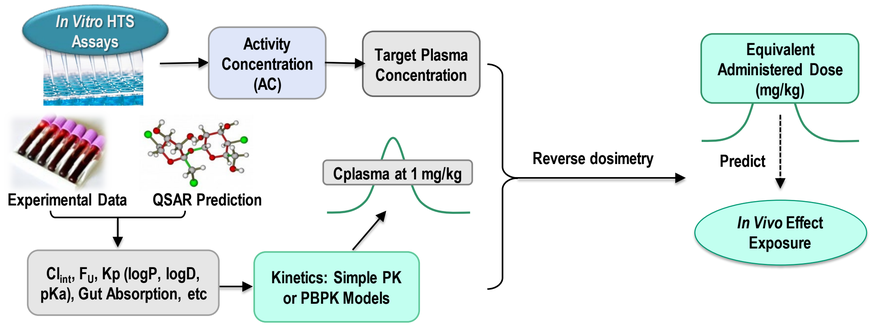


Command line



Graphical user interface

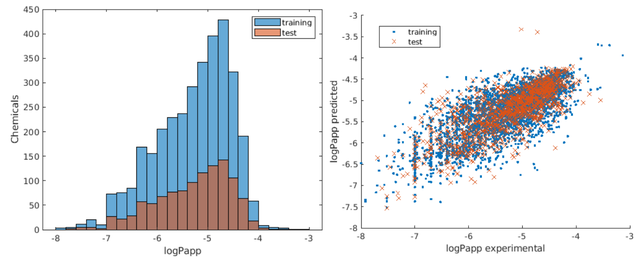
Pharmacokinetic Parameters



Caco2: Permeability

- The Caco2 dataset combined multiple sources yielding a total of 4423 unique chemicals after curation, deduplication, and outlier removal.

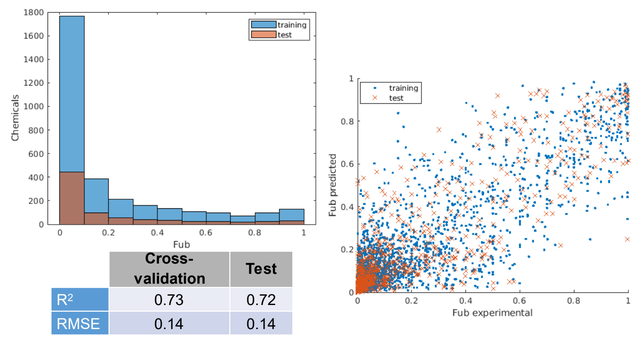
	Cross-validation	Test
R^2	0.57	0.56
RMSE	0.49	0.49



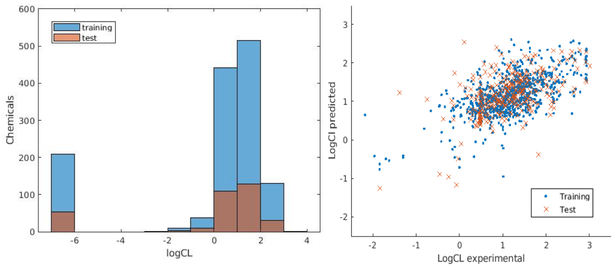
F_U: Fraction unbound

- Both the CL_{int} and F_U OPERA models were recently updated by adding new data from the ChEMBL database to the initial data collected from EPA's high-throughput toxicokinetic (*httk*) R package.
- After several rounds of automated and manual curation to reduce errors, variability and outliers, the F_U dataset totaled 3155 chemicals.

	Cross-validation	Test
R^2	0.73	0.72
RMSE	0.14	0.14



Cl_{int}: Intrinsic clearance

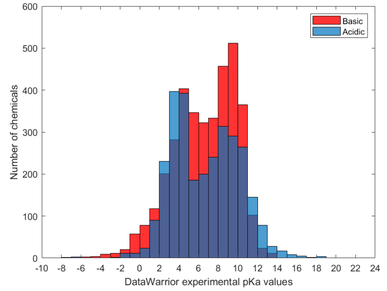


The CL_{int} dataset (1346 chemicals) was modeled in two steps:
- Classification model to separate the cleared from non-cleared chemicals.
- Regression model to predict the CL_{int} value for the cleared chemicals.

	Cross-validation	Test
BA	0.67	0.60
R^2	0.38	0.37
RMSE	0.57	0.57

Physicochemical Parameters

pKa: Acid dissociation constant

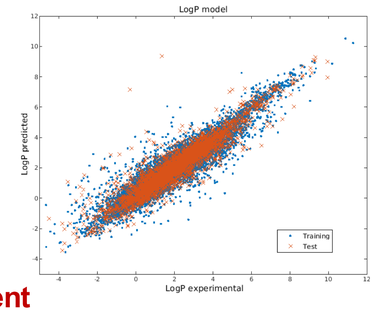


- The acidic (3260 chemicals) and basic (3680 chemicals) datasets were modeled separately.
- First, a weighted k-nearest neighbor classification model predicted whether a chemical is acidic, basic or both with a balanced accuracy of 0.80 and 0.77 for the training and the test sets, respectively.
- Then a support vector machines model predicted the strongest acidic and basic pKa values.

	Acidic pKa		Basic pKa	
	Cross-validation	Test	Cross-validation	Test
R^2	0.64	0.72	0.71	0.78
RMSE	2.02	1.80	1.76	1.53

LogP: Octanol-water partition coefficient

- The OPERA logP model was initially built using a curated dataset from the PHYSPROP database.
- The overall statistics of the model reached an R^2 of 0.86 and an RMSE of 0.78 for the test set.
- The logP model as well as other OPERA models (water solubility, and vapor pressure) have been updated to account for highly investigated groups of chemicals such as polyfluorinated substances (PFAS).



LogD: Distribution coefficient

- LogD is the pH-dependent distribution coefficient that is used to estimate the different relative concentrations of the ionized and non-ionized forms of a chemical at a given pH.
- OPERA uses both pKa and logP predictions to provide logD estimates for ionizable chemicals at pH 5.5 and pH 7.4.
- LogD is estimated using the following formula: $logD_{(pH)} = logP - \log(1 + 10^{(pH-pKa)})$.

References

[1] Mansouri, K. et al. J Cheminform (2018) https://doi.org/10.1186/s13321-018-0263-1
[2] Mansouri, K. et al. SAR & QSAR in Env. Res. (2016) https://doi.org/10.1080/1062936X.2016.1253611
[3] Williams A.J. et al. J Cheminform (2017) https://doi.org/10.1186/s13321-017-0247-6
[4] JRC QSAR Model Database https://qsardb.jrc.ec.europa.eu/qmrf/endpoint
[5] Mansouri, K. et al. EHP (2016) https://doi.org/10.1289/ehp.1510267
[6] Mansouri, K. et al. J Cheminform (2019) https://doi.org/10.1186/s13321-019-0384-1
[7] Mansouri, K. et al. EHP (2020) https://doi.org/10.1289/EHP5580
[8] Kleinstreuer, N. et al. Comp Tox (2018) https://doi.org/10.1016/j.comtox.2018.08.002
[9] Mansouri, K et al. EHP "CATMoS manuscript" (2021) https://doi.org/10.1289/EHP8495

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.