

Open-source QSAR-ready chemical structure standardization workflow

K Mansouri¹, C Grulke², R Judson², A Richard², A Williams², and N Kleinstreuer¹

Abstract 229

¹NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; ²EPA/CCTE, RTP, NC, USA

This work does not necessarily represent the views or policy of the US EPA. Any mention of tradenames does not constitute endorsement

OBJECTIVES

- Provide a free and open-source tool to process chemical structures associated with experimental data for building robust QSAR models, especially from public domain sources
- Define a consistent structure standardization set of rules to apply prior to making predictions on existing QSAR models
- Improve the consistency and accuracy of QSAR models and facilitate their applicability domain assessment

APPROACH

- Use only free and open-source tools: KNIME environment
- Provide a flexible workflow generating consistent QSAR-ready structures from different file formats
- Involve experts in the field to reach consensus about certain arbitrary decisions and choices for molecular rendering
- Share the workflow online in different formats to accommodate different uses.

MAIN RESULTS

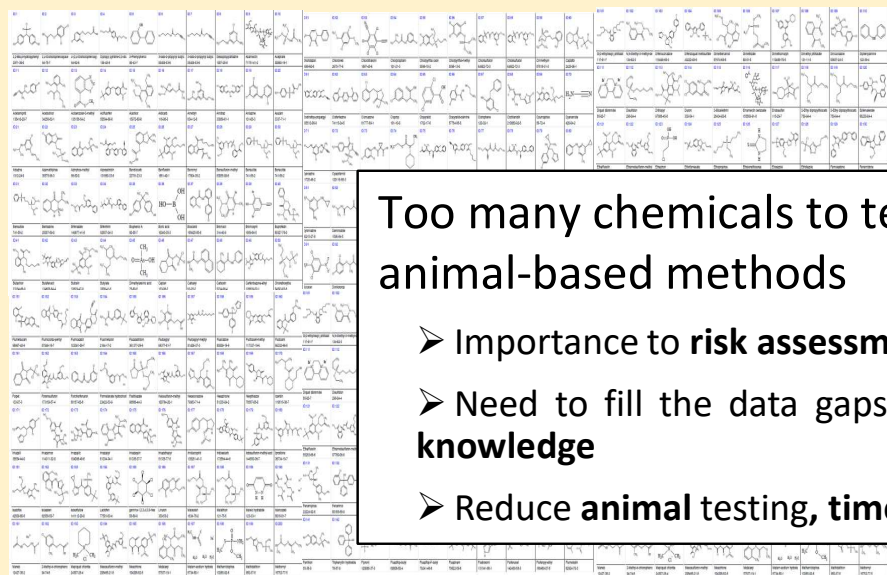
- The initial workflow was designed and built in KNIME desktop version
- The workflow was then modified to provide different input-output file formats and variables for command line use
- The modified version was implemented in OPERA suite of QSAR models and deployed to NIEHS KNIME server and dockerized

IMPACT

- The resulting QSAR-ready standardization workflow was used in three international collaborative modeling projects
- The command line version was implemented in OPERA and used to provide all OPERA predictions that are available on the EPA Comptox Chemicals dashboard and NTP ICE
- The standardized structures are also being used for other purposes such as similarity search, metabolite predictions, and mass-spectrometry analysis.

Open-source QSAR-ready chemical structure standardization workflow

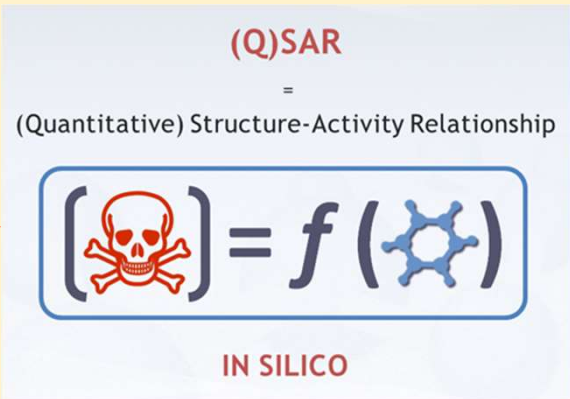
OBJECTIVES



Too many chemicals to test with standard animal-based methods

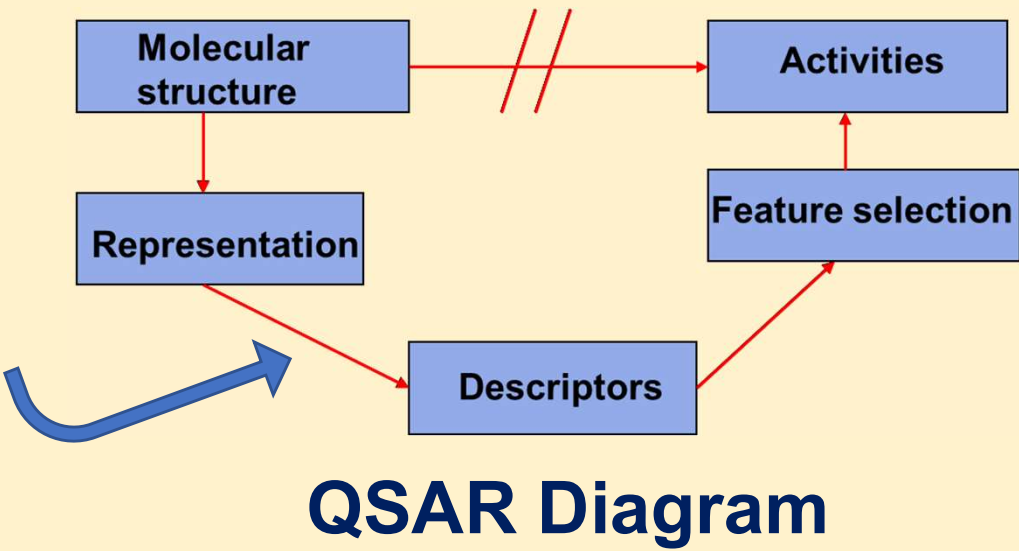
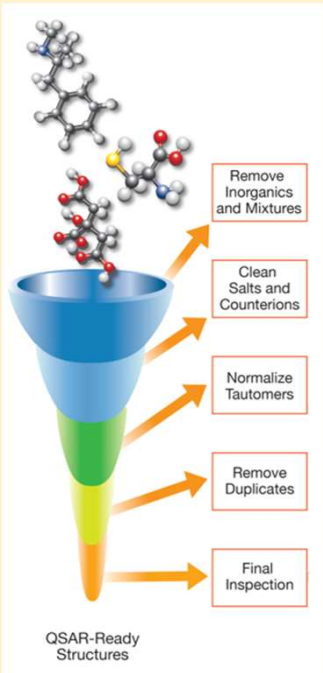
- Importance to **risk assessment** of chemicals
- Need to fill the data gaps and bridge the **lack of knowledge**
- Reduce **animal testing, time and costs**

Alternative



Aim of the workflow:

- Combine different procedures and ideas for structure standardization
- Minimize the issues faced during descriptor calculation and correct or remove inadequate structures
- Increase accuracy and consistency of prediction
- Produce a flexible free and open-source workflow to be shared

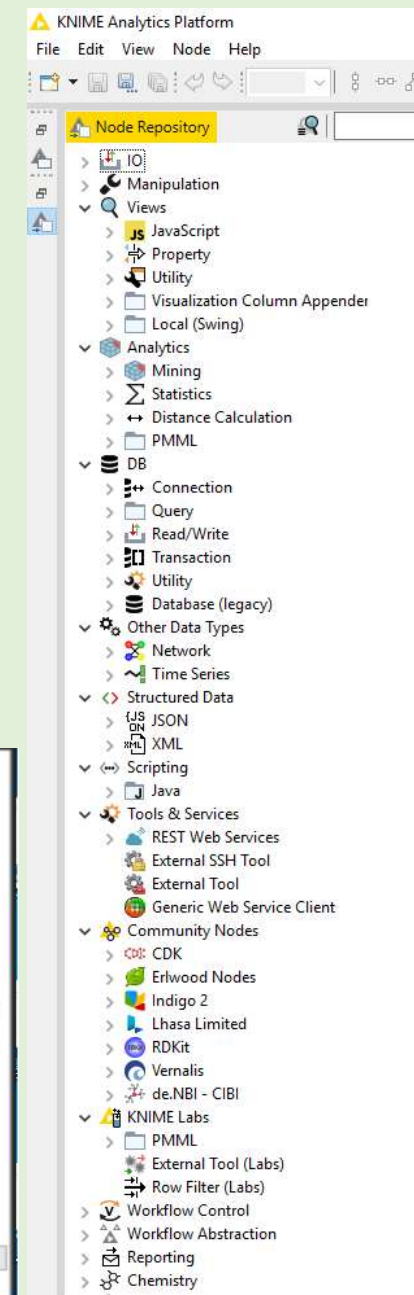
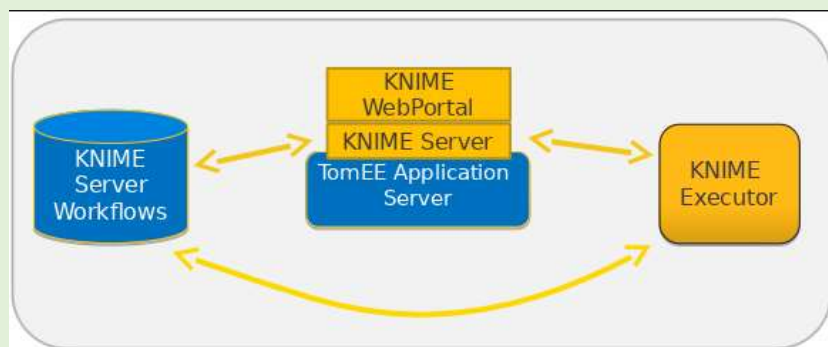
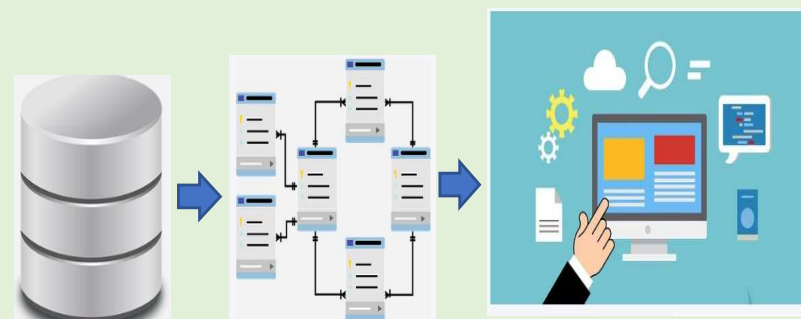


Open-source QSAR-ready chemical structure standardization workflow

APPROACH

KNIME (Konstanz information miner)

- Free (desktop version) and open source
- Provides tools for different fields (scientific, computational, data-mining, text mining, statistics, image processing...)
- Does not require any programming skills
- Results visualization at any step
- Integrates and executes different programming languages
- Includes large number of software tools
- Allows use of commercial/proprietary tools
- Facilitates sharing and collaboration
- Large community developing new tools
- Technical support and user groups/meetings

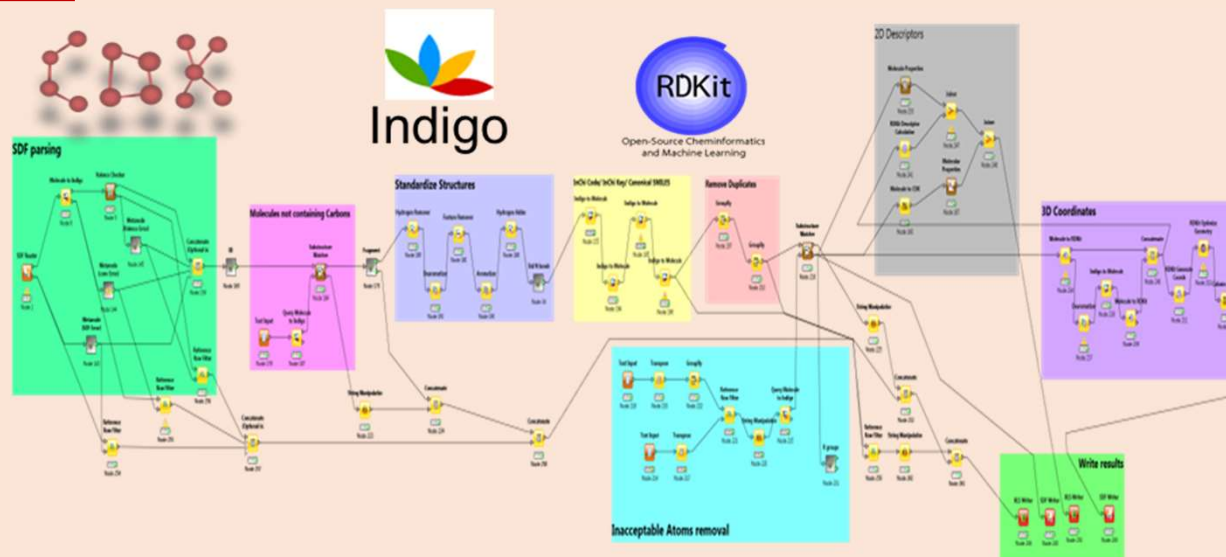


Open-source QSAR-ready chemical structure standardization workflow

MAIN RESULTS

Versions by use

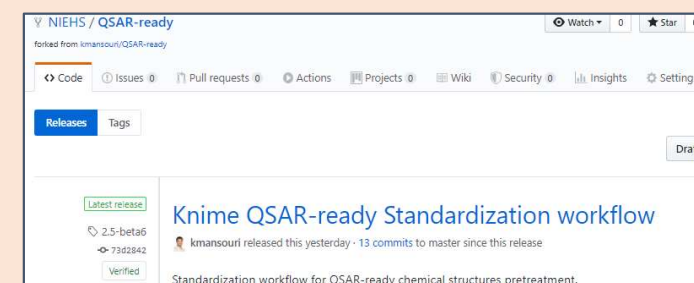
KNIME Standalone



Public availability

Github

<https://github.com/NIEHS/QSAR-ready/releases>



KNIME Hub:

hub.knime.com/kmansouri/spaces/Public/latest/QSAR-ready_2.5.6



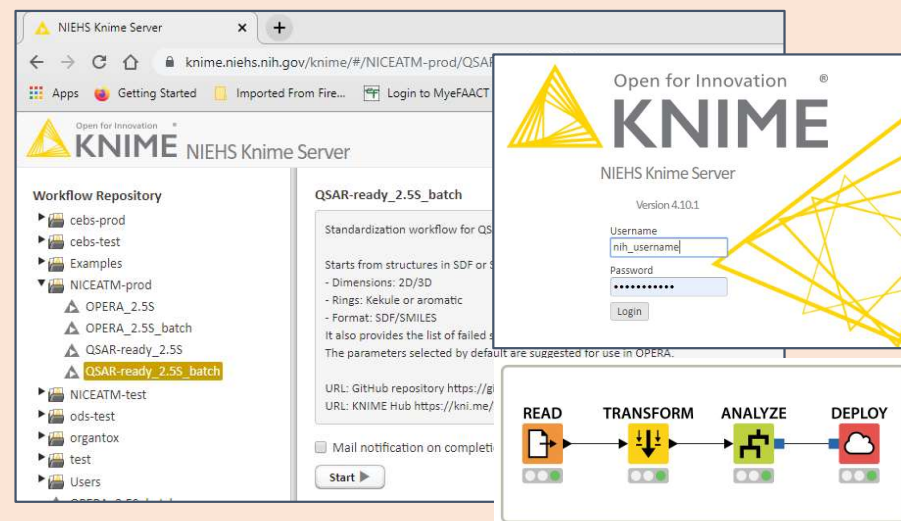
NIEHS KNIME server

KNIME Webportal

knime.niehs.nih.gov/knime/#/NICEATM-prod/QSAR-ready_2.5S_batch

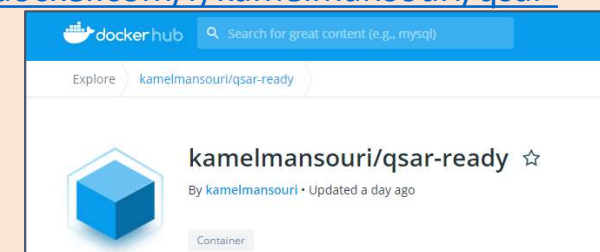
KNIME REST API

https://ehsntpvlp03.niehs.nih.gov:8443/knime/rest/v4/repository/NICEATM-prod/QSAR-ready_2.5S_batch:openapi?showInUI=true



Docker hub

<https://hub.docker.com/r/kamelmansouri/qsar-ready/tags>



Open-source QSAR-ready chemical structure standardization workflow

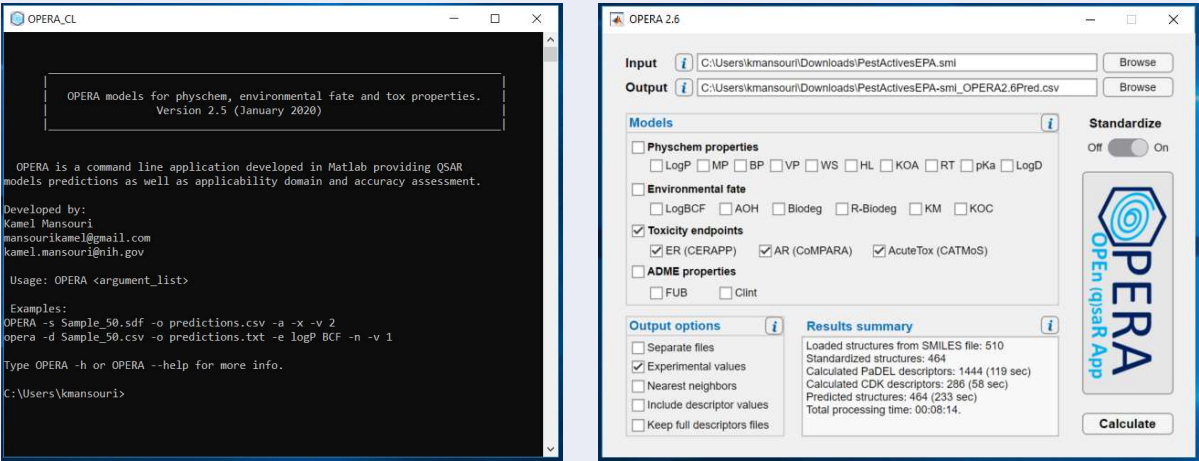
Implementation in OPERA

OPERA suite of QSAR models:

- Free, open-source, and open-data
- Command line and GUI
- Single chemical and batch mode
- Windows OS and Linux
- Embeddable wrapper libraries in Java, C, C++, and Python

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

IMPACT



EPA Dashboard: comptox.epa.gov/dashboard/

EPA United States Environmental Protection Agency

Chemistry Dashboard

Bisphenol A
80-05-7 | DTXSID7020182

Model Results
Predicted vs. Global mean
Lower mean
Contrast

875 Thousand Chemicals

Chemicals Product/Use Categories Assay/Gene

Search for chemical by systematic name, synonym, CAS number

Identifier substring search

See what people are saying, read the dashboard
Cite the Dashboard Publication click here

Batch Search

Step 1 Step 2 Step 3 Step 4 Step 5: Click "Download"

Please enter one identifier per line

Select Input Type(s)

- ☒ Identifiers
 - ☒ Chemical Name
 - ☒ CASRN
 - ☒ InChIKey
 - ☒ DDTTox Substance ID
 - ☒ InChIKey Skeleton
 - ☒ MS-Ready Formula
 - ☒ Exact Formula
 - ☒ Monoisotopic Mass

Enter Identifiers to

Select Output Format:

Excel

Download

Customize Results

- ☒ Select All
- ☒ Select All in Lists
- ☒ Chemical Identifiers
 - ☒ DTXCID
 - ☒ Chemical Name
 - ☒ CAS-RN
 - ☒ InChIKey
 - ☒ IUPAC Name
- ☒ Structures
 - ☒ Mol File
 - ☒ InChI String
 - ☒ MS-Ready SMILES
 - ☒ QSAR-Ready SMILES
- ☒ Intrinsic And Predicted Properties

Presence in Lists:

- ☒ DTXCID
- ☒ CAS-RN
- ☒ InChIKey
- ☒ IUPAC Name
- ☒ Structures
- ☒ InChI String
- ☒ MS-Ready SMILES
- ☒ QSAR-Ready SMILES
- ☒ Intrinsic And Predicted Properties

NTP ICE: ice.ntp.niehs.nih.gov/

National Toxicology Program
U.S. Department of Health and Human Services

ICE v3.0.2 Release

ICE updates include:

- CHTS Summary view to include QC-Omit and Flag-Omit.
- Minor bug fixes.

Learn about ICE updates

UPDATES

Search Data Chemical Characterization INIVE

Acetaminophen 103-90-2 DTXSID2020008 1825

Endosulfen 115-29-7 DTXSID1020560 2.26

3-Phenylprop-2-enal 104-55-2 DTXSID1024835 2568

Acrolein 107-02-8 DTXSID9020023 40

Linalool 78-70-6 DTXSID7025502 2097

Inactive Inactive Inactive Inactive Inactive Inactive

Assay Input

Select Assays

Assay

- ☒ CERAPP, ER Binding
- ☒ CERAPP, ER Antagonist
- ☒ CERAPP, ER Agonist
- ☒ CoMPARA, AR Binding
- ☒ CoMPARA, AR Antagonist
- ☒ CoMPARA, AR Agonist
- ☒ CATMoS, Rat Acute Oral Toxicity