Open-source QSAR-ready chemical structure standardization workflow

K Mansouri[1], C Grulke[2], R Judson[2], A Richard[2], A Williams[2], and N Kleinstreuer[1]

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

Abstract 229

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

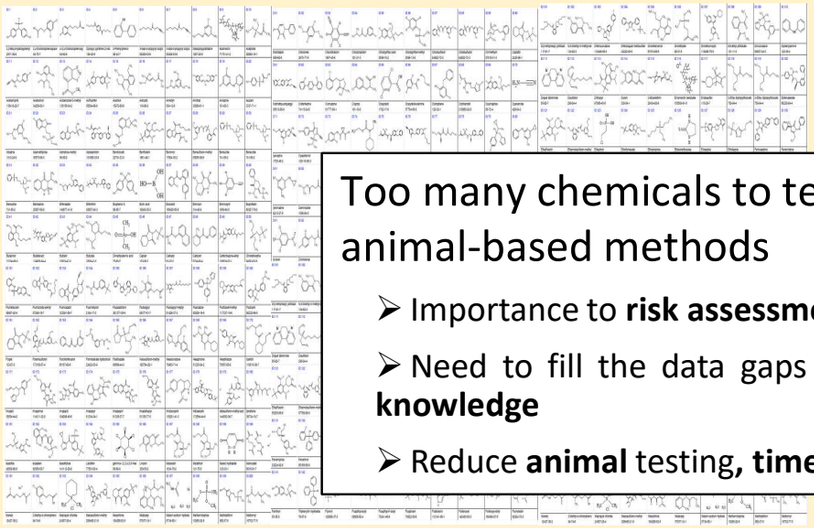
- The initial workflow was designed and built in KNIME desktop version
- The workflow was then modified to provide different input-output file formats and variables for command line use
- The modified version was implemented in OPERA suite of QSAR models and deployed to NIEHS KNIME server and dockerized

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

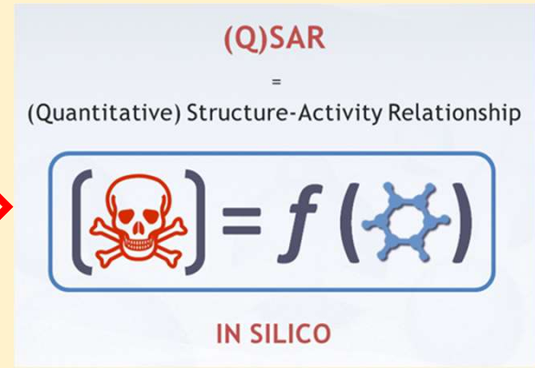
OBJECTIVES



Too many chemicals to test with standard animal-based methods

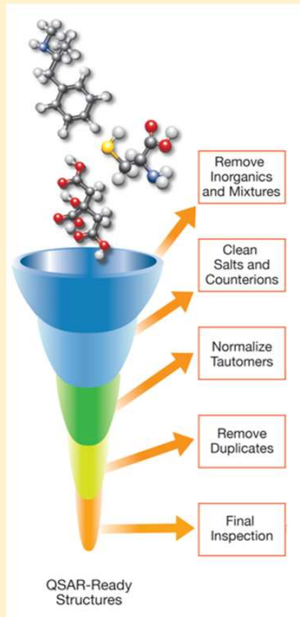
- Importance to **risk assessment** of chemicals
- Need to fill the data gaps and bridge the **lack of knowledge**
- Reduce **animal** testing**, time** and **costs**

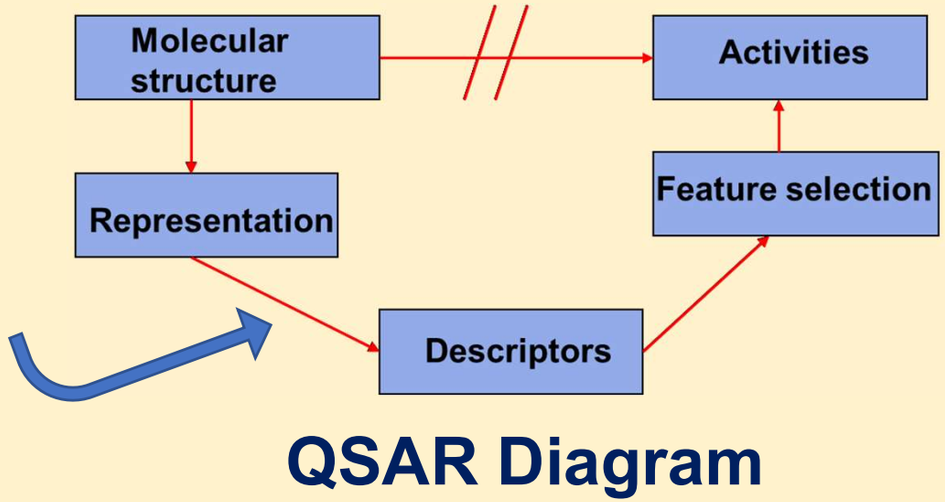
Alternative



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





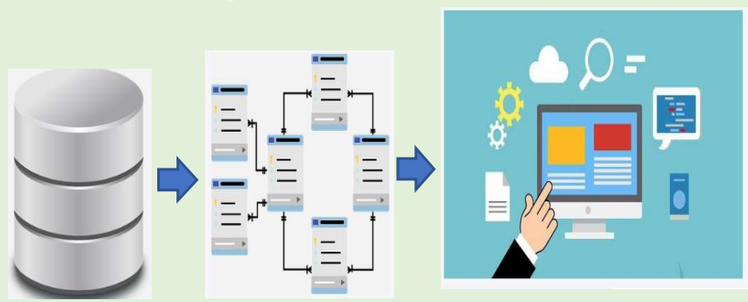
QSAR Diagram

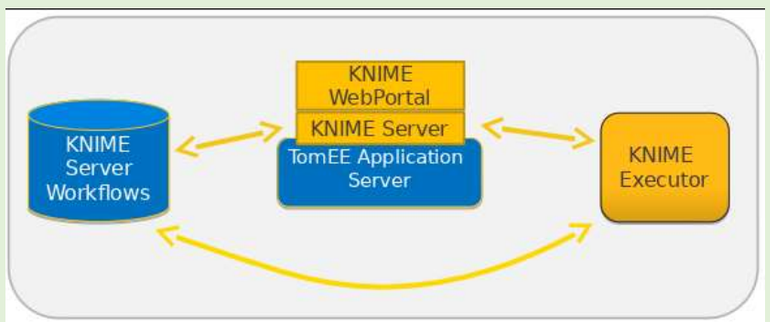
Open-source QSAR-ready chemical structure standardization workflow

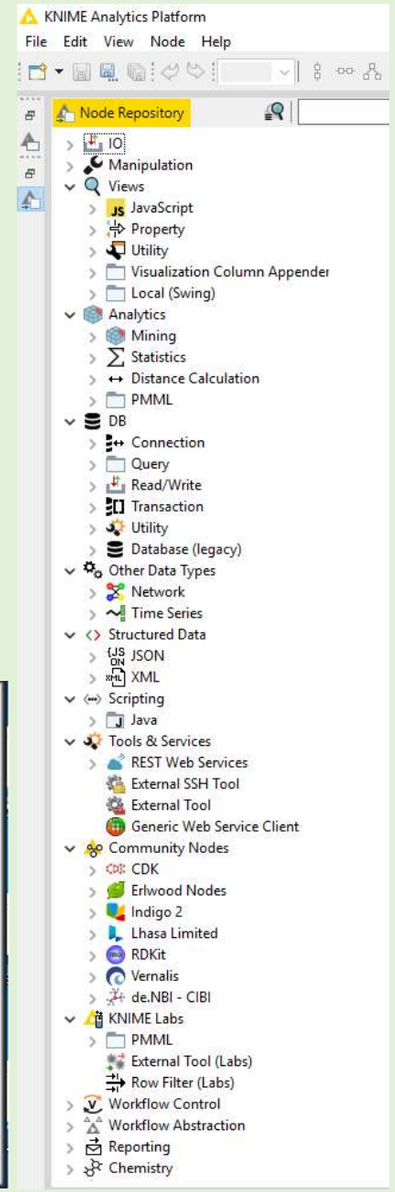
APPROACH

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





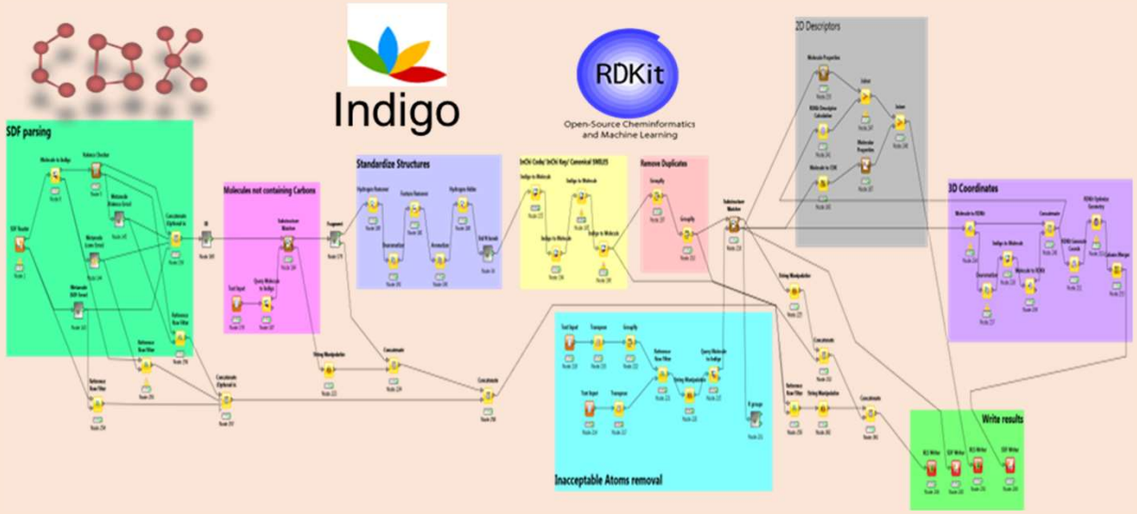


Open-source QSAR-ready chemical structure standardization workflow

MAIN RESULTS

Versions by use

KNIME Standalone



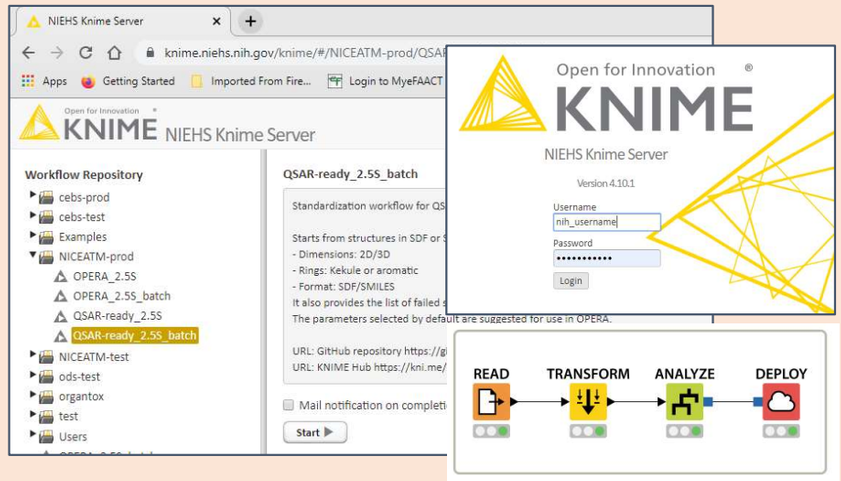
NIEHS KNIME server

KNIME Webportal

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

KNIME REST API

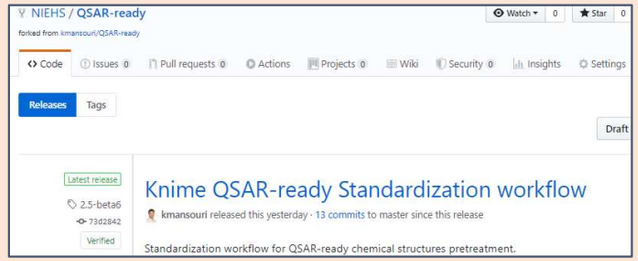
https://ehsntpvlp03.niehs.nih.gov:8443/knime/rest/v4/repository/NICEATM-prod/QSAR-ready_2.5S_batch:openapi?showInUI=true



Public availability

Github

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

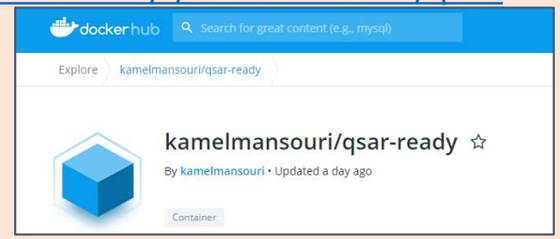


KNIME Hub:

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

Docker hub

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



Open-source QSAR-ready chemical structure standardization workflow

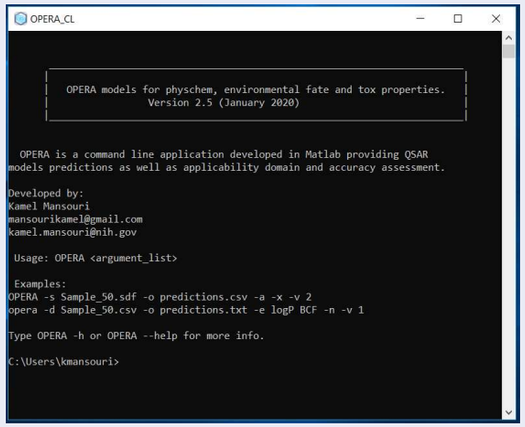
Implementation in OPERA

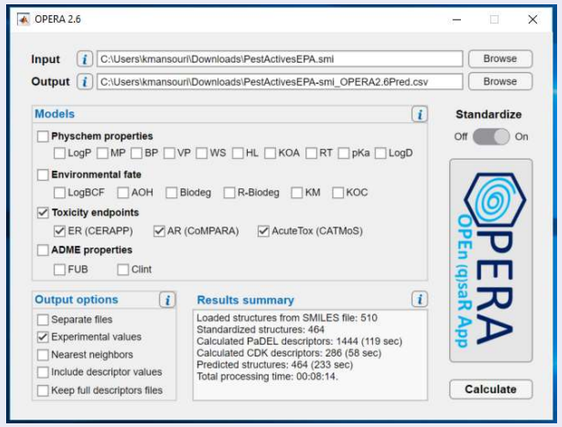
OPERA suite of QSAR models:

- 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

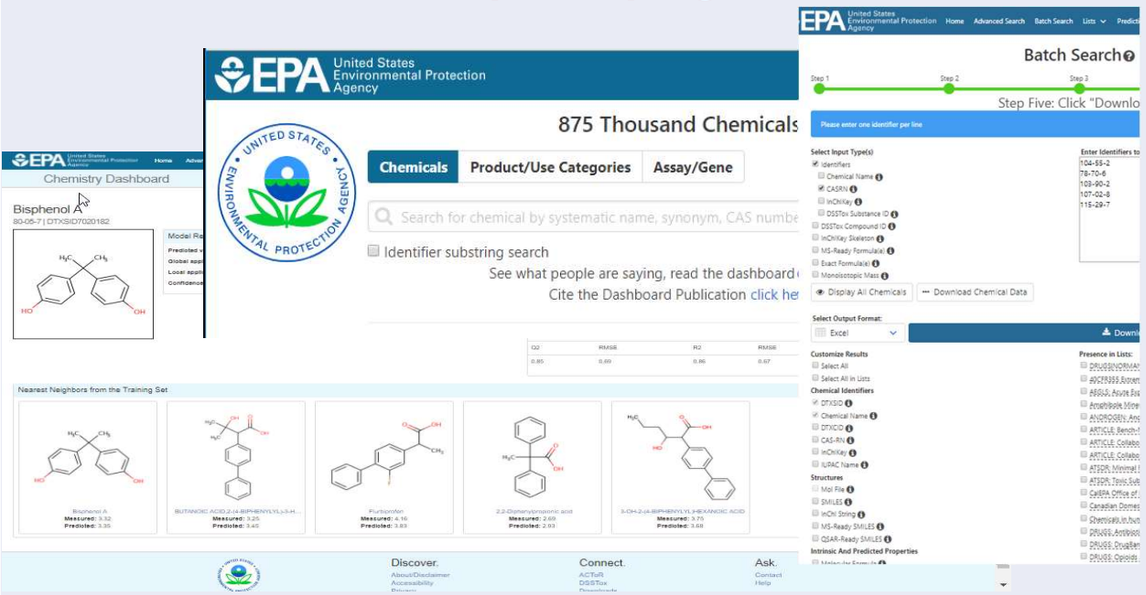
https://github.com/NIEHS/OPERA

IMPACT





EPA Dashboard: comptox.epa.gov/dashboard/



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