



The EPA Online Prediction Physicochemical Prediction Platform to Support Environmental Scientists

ANTONY WILLIAMS

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:
tony27587@gmail.com

DATE RECEIVED:
September 06, 2016

© Williams This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



This poster was presented at the American Chemical Society in Philadelphia in August 2016 at the Sci-Mix gathering and at the ENVR section on Wednesday.

SESSION: Sci-Mix

SESSION TIME: August 22, 2016 from 8:00 PM to 10:00 PM

and

SESSION TIME: Wednesday, August, 24, 2016, 6:00 PM – 8:00 PM

ROOM & LOCATION: Hall D – Pennsylvania Convention Center

Poster Title: The EPA Online Prediction Physicochemical Prediction Platform to Support Environmental Scientists

As part of our efforts to develop a public platform to provide access to predictive models we have attempted to disentangle the influence of the quality versus quantity of data available to develop and validate QSAR models. Using a thorough manual review of the data underlying the well-known EPI Suite software, we developed automated processes for the validation of the data using a KNIME workflow. This includes: approaches to validate different chemical structure representations (e.g. molfile and SMILES), identifiers (chemical names and registry numbers), and methods to standardize the data into QSAR-consumable formats for modeling. Our efforts to quantify and segregate data into various quality categories has allowed us to thoroughly investigate the resulting models developed from these data slices, as well as allowing us to examine whether or not efforts into the development of large high-quality datasets has the expected pay-off in terms of prediction performance. Machine-learning approaches have been applied to create a series of models that have been used to generate predicted physicochemical and environmental parameters for over 700,000 chemicals. These data are available online via the EPA's iCSS Chemistry Dashboard. *This abstract does not reflect U.S. EPA policy.*

The EPA Online Prediction Physicochemical Prediction Platform to Support Environmental Scientists from **US Environmental Protection Agency (EPA)**