American Chemical Society Fall meeting, Aug. 21-25, Philadelphia, Penn.

POSTER: ENVR session posters: Computational Chemistry & Toxicology in Chemical Discovery & Assessment (QSARs)-Poster

Session Outline: Testing chemicals for performance/efficacy, toxicologic effects, and environmental fate is expensive and time consuming. Decisions to conduct such studies often involve cost/benefit evaluations leading to limitations on the data that can be developed. In addition, environmental professionals in recent years have been trying to reduce the numbers of test animals used in toxicologic studies, for ethical and cost reasons. Elucidating the toxicologic mechanism by which a chemical acts with experimental data can be difficult, expensive, and ultimately, equivocal. Finally, scientists often have to predict risks based on little or no data on chemicals and their metabolites in the context of the environmental statutes Toxic Substances Control Act (TSCA) and Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). The goal of computational chemistry and toxicology is to meet these needs, at least partially. This general approach can involve the use of quantum chemistry to model toxicologic initiating events at the molecular level. Quantitative structure activity relationships (QSARs) can be used to predict environmental fate and adverse effects of a chemical. Thus the purpose of this symposium is to present papers on accomplishments and challenges in this field. A related goal is to consider where this prediction methodology will be in five to ten years.

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

Authors: Antony Williams[†], Kamel Mansouri[††], et al (abstract will be updated later)

[†] National Center for Computational Toxicology, Office of Research & Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

[††] ORISE Fellow, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

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.*