Integration of different data gap filling techniques to facilitate assessment of polychlorinated biphenyls: A proof of principle case study

Prachi Pradeep[1], Laura Macaulay[2], Geniece Lehmann[2], Richard S Judson[3], Grace Patlewicz[3]

[1]ORISE, National Center for Computational Toxicology, RTP, NC; [2]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, RTP, NC
[3]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, RTP, NC

pradeep.prachi@epa.gov

Data gap filling techniques are commonly used to predict hazard in the absence of empirical data. The most established techniques are read-across, trend analysis and quantitative structure-activity relationships (QSARs). Toxic equivalency factors (TEFs) are less frequently used data gap filling techniques which are applied to estimate relative potencies for mixtures of chemicals that contribute to an adverse outcome through a common biological target. For example, The TEF approach has been used for dioxin-like effects comparing individual chemical activity to that of the most toxic dioxin: 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. The aim of this case study was to determine whether integration of two data gap filling techniques: QSARs and TEFs improved the predictive outcome for the assessment of a set of polychlorinated biphenyl (PCB) congeners and their mixtures. PCBs are associated with many different adverse effects, including their potential for neurotoxicity, which is the endpoint of interest in this study. The dataset comprised 209 PCB congeners, out of which 87 altered in vitro Ca(2+) homeostasis from which neurotoxic equivalency values (NEQs) were derived. The preliminary objective of this case study was to develop a QSAR model to predict NEQ values for the 122 untested PCB congeners. A decision tree model was developed using the number of position specific chlorine substitutions on the biphenyl scaffold as a fingerprint descriptor. Three different positional combinations were explored on the basis of equivalence between ortho, meta and para positions. Five different decision trees were developed on the basis of restrictions on tree growth. The training dataset of 87 tested PCBs was evaluated using 5-fold cross validation and leave-one-out (LOO) internal validation to ultimately predict NEQ values for the 122 untested PCBs. The evaluation statistics of the “best” decision tree model resulted in LOOCV RMSE: 0.29, 5-fold CV test RMSE: 0.34, and R^2: 0.79. The results demonstrate the utility of using the TEF approach as an alternative data gap filling technique.

This abstract does not necessarily represent U.S. EPA policy.