Prediction of Cytochrome P450 Profiles of Environmental Chemicals with QSAR Models Built from Drug-like Molecules

Hongmao Sun,[1] Henrike Veith,[1] Menghang Xia,[1] Christopher P. Austin,[1] Raymond R. Tice,[2] Robert J. Kavlock,[3] Ruili Huang[1]

[1]National Institutes of Health Chemical Genomics Center, Rockville, MD

[2] National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC

[3] National Center for Computational Toxicology, Office of Research and Development, US EPA, Research Triangle Park, NC

The human cytochrome P450 (CYP450) enzyme family is involved in the biotransformation of many environmental chemicals. As part of the U.S. Tox21 effort, we profiled the CYP450 activity of ~2800 chemicals predominantly of environmental concern against CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4 isoforms in a quantitative high throughput screening (qHTS) format. Five support vector machines (SVM) models built from a large dataset consisting of over 17,000 drug-like compounds were challenged to predict the CYP450 activities of the Tox21 Phase I compound collection. Although a large fraction of the test compounds fall outside of the applicability domain (AD) of the model, as measured by *k*-nearest neighbor (*k*-NN) similarities, the predictions were accurate for CYP1A2, CYP2C9, and CYP3A4 ioszymes with area under the receiver operator characteristic curves (AUC-ROC) ranging between 0.83 and 0.85. However, the CYP2C19 and CYP2D6 models had lower predictive power with AUC-ROC of 0.76 due to larger variations in the CYP2C19 training data and imbalanced training (13.8% positives) and test data (9.9% positives) for the CYP2D6 model. Several different rebalancing strategies were applied to the CYP2D6 dataset, and the consensus of the random under-sampling method outperformed other re-sampling methods. Atom type-based models achieved better predictive accuracies over all five CYP450 isoforms than those based on ECFP_6 and MOE 2D descriptors. Our results demonstrate that decomposing molecules into atom types enhanced the coverage of AD, thus the models can be used to predict the ability of non-drug like compounds to interact with these CYPs. Supported by NIEHS Interagency Agreement Y3-ES-7020-01 and EPA Interagency Agreement Y3-HG-7026-03.

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