A Systematic Evaluation of Analogs for the Read-across Prediction of Estrogenicity

Prachi Pradeep[1,2], Kamel Mansouri[1,2], Grace Patlewicz[1], Richard Judson[1]

[1]NCCT, ORD, US EPA, North Carolina, USA
[2]ORISE, Oak Ridge, Tennessee, USA

Read-across is a data gap filling technique widely used within category and analog approaches to predict a biological property for a target data-poor chemical using known information from similar (source analog) chemical(s). Potential source analogs are typically identified based on structural similarity. Although much guidance has been published for read-across, practical guiding principles for the identification and evaluation of the scientific validity of source analogs, which is a critical step in deriving a robust read-across prediction, remains largely lacking. This case study explores the extent to which 3 structure descriptor sets (Pubchem, Chemotyper and MoSS) and their combinations are able to identify valid analogs for reading across Estrogen Receptor (ER) activity for a specific class of chemicals: hindered phenols. For each target chemical, two sets of analogs (hindered and non-hindered phenols) were selected using each descriptor set with two cut-offs: (1). Minimum Tanimoto similarity (range 0.1 - 0.9), and (2). Closest *N* analogs (range 1 - 10). Each target-analog pair was then evaluated for its agreement with measured ER binding and agonism. Subsequently, the analogs were filtered using physchem properties ($LogK_{ow}$ & Molecular Volume) and the resultant agreement between each target-analog pair was evaluated. The data set comprised 462 hindered phenols and 296 non-hindered phenols. The results demonstrate that: (1) The concordance in ER activity rises with increasing similarity, (2) none of the 3 descriptor sets are clearly superior to the others for analog identification for ER read-across, (3) selecting hindered versus non-hindered phenols as analogs does not significantly improve concordance in ER activity, and (4) filtering of analogs using physchem properties improves overall concordance. As an example, selecting hindered phenols as analogs with a similarity cut-off of 0.9, the maximum concordance observed is 76% (Chemotyper+MoSS) for ER binding and 93% (Pubchem+MoSS) for agonism. Filtering with physchem properties increases these values to 78% (Chemotyper+Moss) for ER binding and 95% (Chemotyper alone) for agonism. This case study demonstrates how biologically-relevant chemical descriptors can be used to identify valid analogs for read-across.

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