In-silico study of ToxCast GPCR assays by Quantitative Structure-Activity Relationships (QSARs) modeling.

Mansouri K., Sipes N.S., Judson R.S.

National Center for Computational Toxicology, ORD/U.S. EPA/ RTP/ NC/ USA.

Several thousand chemicals were tested in 700 toxicity-related *in-vitro* HTS bioassays through the EPA's ToxCast and Tox21 projects. This chemical set only covers a portion of the chemical space of interest for environmental exposure, leading to a need to fill data gaps with alternative methods. A cost effective and reliable approach to fulfill this task is to build Quantitative Structure-Activity Relationships (QSARs). In this work, a subset of 1800 ToxCast chemicals was used to build QSAR models for multiple ToxCast assays to predict activity of chemicals in a larger environmental database of ~30K structures. The initial molecular targets for this project were a set of 18 G-Protein Coupled Receptor (GPCR) assays. These assays are part of the aminergic category which was among the most active within the biochemical assays. The QSAR predictions were based on two levels; the first was a classification into active/non-active chemicals; then regression models were built to predict the AC50 potency values of the bioassays for the active chemicals. Different software packages were used to calculate constitutional, topological and fingerprinted molecular descriptors based on two-dimensional structures. Then several classification and regression model-fitting methods including PLS-DA, SVM, MLR, PLS and *k*NN were tested. The overall approach also included variable selection techniques such as Genetic Algorithms that were applied in order to select the most predictive molecular descriptors for each assay. The models were evaluated using *n*-fold cross-validation and forward validation on a held-out subset of the initial data. Finally, the applicability domains of the models were defined. Using PLS-DA for the human histamine H1 GPCR assay, a classification accuracy of 94% was reached with a non-error rate of 89% in fitting and 80% with 5-fold cross-validation with only 2 latent variables. This work shows the promise of using ToxCast in vitro data to develop structure-based models for use in predicting target activity across the diverse space of environmental chemicals. *This abstract does not necessarily reflect U.S. EPA policy.*