

QSAR Modeling for the Predictions of Androgen Receptor Pathway Activity

Q Zang¹, N Kleinstreuer¹, D Allen¹, W Casey², R Judson³

¹ILS/NICEATM, RTP, NC, USA; ²NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; ³EPA/ORD/NCCT, RTP, NC, USA

Fast and cost-effective approaches are needed to evaluate the potential of thousands of man-made chemicals to disrupt the endocrine system, including interaction with the androgen receptor (AR) pathway. The Tox21 and ToxCast programs have tested ~1800 chemicals in a broad panel of *in vitro* high-throughput screening (HTS) assays. Nine assays that map to the AR pathway have been integrated into a computational model to identify substances with potential androgenic/anti-androgenic activity *in vivo*. The present work uses the HTS dataset, the associated computational model output, and machine learning methods to develop quantitative structure–activity relationship (QSAR) models to predict AR antagonism. Although most ToxCast chemicals (1517) were predicted to be inactive against the pathway, 225 were predicted to have some antagonist activity. QSAR classification models were built to relate the molecular structures of chemicals to predicted AR activities using linear discriminant analysis, classification and regression trees, and support vector machines (SVM) with 51 molecular descriptors from QikProp and 6293 structural fingerprints as potential variables. A random forest (RF) feature selection method was used to extract structural features most relevant to AR activity. A training set of 1161 chemicals was used to derive and optimize the binary classification models. A test set of 581 chemicals was used to validate the performance of each model for overall accuracy, sensitivity, specificity and G-mean. In addition to binary classification, the models predicted potency of the 225 active compounds using multiple linear regression and partial least squares regression. The best performing model was obtained using SVM in combination with a subset of descriptors identified via the RF algorithm. This model was then used to make predictions for a broader chemical universe, predicting that 20.6% (6475/31428) of these chemicals may have AR antagonist activity. This initial result is certainly an overestimate that is confounded by very weak activity or cytotoxicity, and further refinement of the model should improve specificity. *This project was funded in whole or in part with Federal funds from the NIEHS, NIH under Contract No.HHSN273201500010C.*