QSAR Modeling to Predict Androgen Receptor Pathway Activity

Q Zang[1], D Allen[1], W Casey[2], R Judson[3], N Kleinstreuer[1]

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

The Tox21 and ToxCast programs have tested ~1800 chemicals in a broad panel of in vitro high-throughput screening (HTS) assays. Nine HTS assays that map to the androgen receptor (AR) pathway were integrated into a computational model to identify substances with potential androgenic/anti-androgenic activity. Here, we used 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 the HTS data predicted that most ToxCast chemicals would be inactive against the pathway, 225 were predicted to have some antagonist activity. QSAR classification models were trained on these data and built to predict AR antagonist activity using linear discriminant analysis, random forest, and support vector machines. Potential model variables included molecular descriptors from QikProp and MOE and structural fingerprints. Recursive feature elimination was used to extract structural features most relevant to AR activity. A training set of 1243 ToxCast chemicals with AR assay data was used to derive and optimize the binary classification models. A test set of 612 ToxCast chemicals with AR assay data was used to validate the performance of each model. Performance was assessed by overall accuracy, sensitivity, specificity and geometric mean. In addition to binary classification, the models predicted potency of the active compounds using multiple linear regression and partial least squares regression with variables selected by genetic algorithm. The best classification model was 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 work does not reflect EPA policy.* *This project was funded in whole or in part with Federal funds from the NIEHS, NIH under Contract No.HHSN273201500010C.*

Keywords: endocrine, androgens, QSAR, machine learning