QSAR Modeling to Predict Androgen Receptor Pathway Activity

Qingda Zang[1], Nicole Kleinstreuer[1], David Allen[1], Warren Casey[2], Richard Judson[3]

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

Testing for potential endocrine disruption is mandated for thousands of chemicals. We used *in vitro* high-throughput screening assays, an associated computational model output, and machine learning methods to develop quantitative structure–activity relationship (QSAR) models to predict androgen receptor (AR) antagonism. QSAR classification models were built using linear discriminant analysis, classification and regression trees, and support vector machines to relate the molecular structures of chemicals to predicted AR activities. Molecular descriptors from QikProp and structural fingerprints were used as potential variables. A random forest feature selection method was used to extract structural features most relevant to AR activity. Binary classification models were derived and optimized using a training set of 1161 chemicals, and a test set of 581 chemicals was used to validate the performance of each model. In addition to binary classification, we built models to predict potency of the active compounds using multiple linear regression and partial least squares regression. The best binary classification model was used to predict activities for a broader chemical universe. An initial prediction that 20.6% (6475/31428) of these chemicals may have AR antagonist activity is certainly an overestimate that is confounded by very weak activity or cytotoxicity. 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. This abstract does not reflect EPA policy.*

Keywords:

1. High-throughput screening (HTS)
2. Computational toxicology
3. ToxCast/Tox21