Toxico-Cheminformatics and QSPR Modeling of the Carcinogenic Potency Database

Authors: Kun Wang[1], Ann Richard[2], Ivan Rusyn[3], Alexander Tropsha[1]
[1]Division of Medicinal Chemistry and Natural Products, University of North Carolina, Chapel Hill, NC 27599
[2]National Center for Computational Toxicology, US Environmental Protection Agency, Research Triangle Park, NC 27711
[3]Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC 27599

Keywords: toxico-chemoinformatics, QSPR, carcinogenicity, CPDB, kNN

We report on the development of a tiered, confirmatory scheme for prediction of chemical carcinogenicity based on QSAR studies of compounds with available mutagenic and carcinogenic data. For 693 such compounds from the Carcinogenic Potency Database characterized molecular topological descriptors, k-nearest-neighbor QSAR models were developed for (i) mutagens vs. non-mutagens, (ii) carcinogens vs. non-carcinogens, (iii) genotoxic carcinogens vs. non-genotoxic carcinogens, and (iv) genotoxic carcinogens vs. genotoxic non-carcinogens. Except for case (ii), models with predictive accuracy exceeding 80% each were obtained for training, test and external validation sets. Our results compare favorably with those generated with more common fragment-based approaches. Our method can reduce the rate of false positives/negative and strengthen the prediction performance of the QSAR models. This concept can be generalized and extended to better integrate other types of carcinogenicity characteristics in aiding classification, e.g., tumor sites, TD50 range, multisite, multisex, multispecies tumor incidences, etc.

Point of Contact:
Alexander Tropsha
Professor
UNC-CH/School of Pharmacy
Chapel Hill, NC 27599
919-966-2955
tropsha@email.unc.edu