Impact of Machine Learning in Predictive ADME and Kinase Selectivity

Steven J. Wilkens
Takeda California, Inc. 10410 Science Center Dr. San Diego, CA 92121 USA

Two applications of machine learning techniques are described. The first example involves employing an autonomous model construction technique. This technique chooses appropriate fingerprints and learning methodologies to produce models of maximal predictive quality. For this approach, conventional extended connectivity fingerprints (ECFP) molecular feature descriptors were chosen, as well as open source machine learning software (Weka), to build a high quality model for the prediction of the octanol-water partition coefficient (logD). Our model has maintained a strong correlation since its internal deployment. In total, over 800 molecules have been evaluated since publication of the model. In our second example, using data from an in-house kinase screen composed of 498 compounds that were tested against 73 kinases, we combined Weka machine learning algorithms with Monte Carlo simulations to identify small kinases sets whose selectivity scores (*S*) were predictive of the large panel *S* values. This approach was validated by showing that *S* values for the panel of twelve kinases yielded a high degree of correlation with *S* values obtained from much larger kinase panels ranging from 50 to over 300 kinases generated for 66 compounds.