Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning

Andrew F. Zahrt ,* Jeremy J. Henle,* Brennan T. Rose, Yang Wang, William T. Darrow, Scott E. Denmark.†

Catalyst design in asymmetric reaction development has traditionally been driven by empiricism, wherein experimentalists attempt to qualitatively recognize structural patterns to improve selectivity. Machine learning algorithms and chemoinformatics can potentially accelerate this process by recognizing otherwise inscrutable patterns in large datasets. Herein, we report a computationally guided workflow for chiral catalyst selection using chemoinformatics at every stage of development. Robust molecular descriptors that are agnostic to the catalyst scaffold allow for selection of a universal training set on the basis of steric and electronic properties. This set can be used to train machine learning methods to make highly accurate predictive models over a broad range of selectivity space. Using support vector machines and deep feed-forward neural networks, we demonstrate accurate predictive modeling in the chiral phosphoric acid–catalyzed thiol addition to N-acylimines.