Title: Predicting Highly Enantioselective Catalysts Using Tunable Fragment Descriptors

Pavel Sidorov, Hokkaido University

Short abstract: Asymmetric catalysis is one of the most important techniques in the modern organic synthesis. However, catalyst optimization process typically relies on inductive and qualitative assumption of chemists based on screening data. While machine learning models using calculated 3D structures enable quantitative data evaluation, costly quantum chemical calculations are often required. In contrast, using models based on 2D molecular descriptors such as binary molecular fingerprints is time- and cost-efficient, but their predictive performance remains insufficient. In this work, we design new selective imidodiphosphorimidate (IDPI) catalysts for hydroalkoxylation reactions. We employ a machine learning model based on molecular fragments, which are fine-tuned for asymmetric catalysis and represent cyclic or polyaromatic hydrocarbons, enabling robust and efficient virtual screening. The resulting model can process catalyst structures, reaction substrates and products, as well as reaction condition simultaneously, thus providing a general approach to modeling catalyzed reaction outcome. Moreover, while the model was trained using data with only moderate selectivities, we have successfully performed the virtual screening of previously experimentally untested catalysts and selected and validated experimentally new hits showing higher selectivities in a previously unaddressed transformation.