Reaction-Agnostic Featurization of Bidentate Ligands for Bayesian Ridge Regression of Enantioselectivity

A. A. Schoepfer[1,2], R. Laplaza[1], J. Waser[2]*, C. Corminboeuf[1]*

[1]Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, EPFL, 1015 Lausanne, Switzerland, [2]Laboratory of Catalysis and Organic Synthesis, Institute of Chemical Sciences and Engineering, EPFL, 1015 Lausanne, Switzerland

Chiral ligands are important components in asymmetric homogeneous catalysis, but their synthesis and screening can be both time-consuming and resource-intensive. Data-driven approaches have the potential to reduce the time and resources needed for reaction optimization by more rapidly identifying an ideal catalyst than random screening, but are often non-transferrable across reactions.
Here, we introduce a general featurization strategy for bidentate ligands, coupled with an automated feature selection pipeline and Bayesian ridge regression (BRR) for multivariate linear regression (MLR) modeling. Our approach, which is applicable to any reaction, incorporates electronic, steric, and topological features, such as rigidity/flexibility, branching, geometry, or constitution, and is well-suited for early-stage ligand optimization. With only 19 to 30 data points per dataset, we validate our workflow for the prediction of enantioselectivity in four metal-catalyzed asymmetric reactions. BRR provides uncertainty estimates that can be used in Bayesian optimization (BO) to efficiently explore pools of prospective ligands. We screen 312 chiral bidentate ligands extracted from a crystallographic repository and suggest promising ligand candidates for a challenging asymmetric oxy-alkynylation reaction.

