Enabling late-stage drug diversification by high-throughout experimentation with geometric deep learning.

David Nippa

F. Hoffmann-La Roche AG, Basel, Switzerland
david.nippa@roche.com

Late-stage functionalization (LSF) represents an economical approach for optimizing the properties of drug candidates. However, the chemical complexity of drug molecules often renders LSF challenging. Aiming to address this problem, an LSF platform based on high-throughput experimentation (HTE) and geometric deep learning is established. Our study focuses on late-stage borylation reactions, which provide opportunities for extending structure-activity relationships (SAR) as well as modulation of absorption, distribution, metabolism and extraction (ADME) through consequent broad diversification. Geometric deep learning has shown diverse successful applications to chemistry. Herein, a geometric deep learning platform is introduced that incorporates steric and electronic information to predict reaction outcomes, ideal conditions and regioselectivity. The resulting computational models correctly forecasted the reactivity for 81% of novel substrates. Reaction yields for diverse reaction conditions were predicted with a mean absolute error margin of 4–5%. The regioselectivity of the major products was accurately captured for up to 90% of the cases studied. Applied to 23diverse commercial drug molecules, the platform successfully identified numerous opportunities for structural diversification.