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

David F. Nippa[1,2,†], Kenneth Atz[3,†], Remo Hohler[1], Alex T. Müller[1], Andreas Marx[1], Christian Bartelmus[1], Georg Wuitschik[1], Irene Marzuoli[1], Vera Jost[1], Jens Wolfard[1], Martin Binder[1], Antonia F. Stepan[1], David B. Konrad[2,*], Uwe Grether[1,*], Rainer E. Martin[1,*] & Gisbert Schneider[3,5,*]

[1]Roche Pharma Research and Early Development (pRED), Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, 4070 Basel, Switzerland.

[2]Department of Pharmacy, Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

[3]ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland.

[4]Process Chemistry and Catalysis (PCC), F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, 4070 Basel, Switzerland.

[5]ETH Singapore SEC Ltd, 1 CREATE Way, #06-01 CREATE Tower, Singapore, Singapore.

[†] These authors contributed equally to this work.

email of the main speaker: kenneth.atz@pharma.ethz.ch

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. [1] Aiming to address this problem, an LSF platform based on high-throughput experimentation (HTE) and geometric deep learning is established. [2] 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. [3] 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 23 diverse commercial drug molecules, the platform successfully identified numerous opportunities for structural diversification.

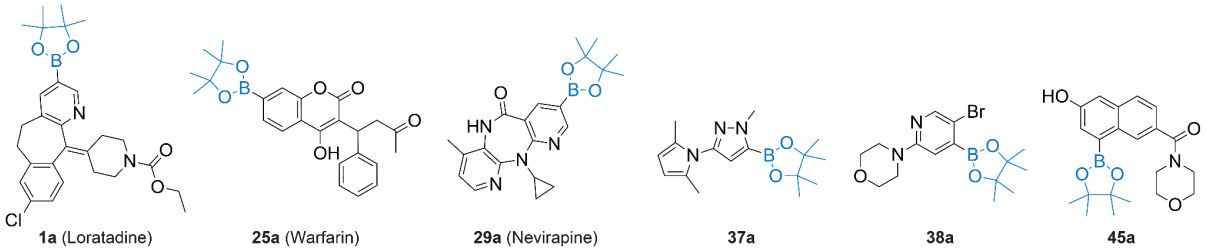

Figure 1: Selected examples of validated borylation opportunities as predicted by machine learning.

[1] D. F. Nippa, *et al*. *Chimia*, **76**, 3, 258-258 (**2022**).
[2] D. F. Nippa, K. Atz, *et al*. *ChemRxiv*, https://doi.org/10.26434/chemrxiv-2022-gkxm (**2022**).
[3] K. Atz, F. Grisoni, G. Schneider, *Nat. Mach. Intell.*, **3**, 1023–1032 (**2021**).