Leveraging historical small molecule data from the Novartis portfolio: From chemical series evolution to predictive modelling

M. Beckers[1], N. Sutrm[1], N. Fechner[1], N. Stiefl[1]

[1]Novartis Pharma AG

In this contribution, we plan to give an overview of our recent activities towards reconstructing the historical landscape of chemical series used in past projects in the Novartis compound archives and how we leveraged this data to build predictive models. In the first part, we will present how we reconstructed the chemical series using clustering methods and show the subsequent statistical analyses of the dataset. These analyses cover the time evolution of structural properties, ADMET and target activities during optimization of the chemical series [1]. Next, we will show how we used this data to build machine learning models for the prediction of small molecule developability. To this end, we further extended the dataset and annotated the terminal development stages the compounds reached. Using these historical milestones together with the recently developed MELLODDY model [2], we built a deep neural network that takes the predicted ADMET profile of a compound as input and returns a number which describes the likelihood of transitioning beyond in vivo PK studies [3]. The resulting score, which we termed bPK score, showed strong performance in discrimination of development candidates from non-development candidates (AUC>0.8). Interestingly, on these internal and public datasets other compound scoring techniques did not show any discriminative performance anymore. We will discuss the retrospective application to MedChem projects and show recent examples of prospective usage.

[1] Beckers, M., Fechner, N. & Stiefl, N. 25 Years of Small-Molecule Optimization at Novartis: A Retrospective Analysis of Chemical Series Evolution. J Chem Inf Model 62, 6002–6021 (2022).
[2] Oldenhof, M. et al. Industry-Scale Orchestrated Federated Learning for Drug Discovery. (2022) doi:10.48550/arxiv.2210.08871.
3] Beckers, M., Sturm, N., Fechner, N. & Stiefl, N. Prediction of small molecule developability using large-scale in silico ADMET models. submitted.