Graph Neural Network Model with Uncertainty Estimation for the Prediction of Xenobiotic Sites of Metabolism

Roxane A. Jacob[1,2,3], Ya Chen[1], Oliver Wieder[1,2], Johannes Kirchmair[1,2]

[1]Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria.
[2]Christian Doppler Laboratory for Molecular Informatics in the Biosciences, Department for Pharmaceutical Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria
[3]Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences (PhaNuSpo), University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

Predicting the metabolically labile atom positions in small organic molecules (i.e. “sites of metabolism”, “SoMs”) is central to the development of safe and efficacious bioactive compounds such as drugs and agrochemicals.[1] State-of-the-art SoM predictors use physicochemical descriptors and/or atom-centred fingerprints in combination with conventional machine learning algorithms such as random forest and/or multilayer perceptrons to rank the atoms of any given small organic molecule according to their metabolic liability.[2] Publicly available labelled data on xenobiotic metabolism is notoriously scarce.[1] Non-proprietary data sets are mostly confined to a limited set of drug-like compounds loosely bound by Lipinski’s rule of five.[4,5] The covered chemical space is too narrow to draw meaningful conclusions concerning the applicability of existing SoM predictors to structurally more diverse compounds such as pesticides. Moreover, the lack of robust uncertainty quantification of the model’s predictions hampers their trustworthiness and acceptance by experimental chemists. To address these challenges, we propose a Graph Neural Network (GNN)-based SoM predictor with epistemic and aleatoric uncertainty quantification. This model, when compared to the established FAME 3 model[2], demonstrates similar performance in terms of accuracy and top-2 success rate. What sets this model apart is its ability to provide an atom-based measure of predictive uncertainty. Specifically, it distinguishes between two types of uncertainty: aleatoric uncertainty, which identifies local chemical environments associated with noisy SoM annotation, and epistemic uncertainty, which highlights unfamiliar chemical structures. We hope that this nuanced analysis will aid experimental chemists in determining when to trust the predictions of SoM models and, ultimately, enhance the acceptance and utility of such tools.

Bibliography:

[1] J. Kirchmair, A. H. Göller, D. Lang, Jens Kunze, B. Testa, I. D. Wilson, R. C. Glen, G. Schneider. Nat. Rev. Drug Discovery 14:6 (2015) 387–404.
[2] M. Šícho, C. Stork, A. Mazzolari, C. De Bruyn Kops, A. Pedretti, B. Testa, G. Vistoli, D. Svozil, J. Kirchmair. JCIM 59:8 (2019) 3400–3412.
[3] N. L. Dang, M. K. Matlock, T. B. Tyler, s. J. Swamidass. JCIM 60:3 (2020) 1146-1164.
[4] J. Zaretzki, M. K. Matlock, S. J. Swamidass. JCIM 53:12 (2013) 3373-3383.
[5] A. Pedretti, A. Mazzolari, G. Vistoli, B. Testa. J. Med. Chem. 61:3 (2018) 19019-1930.