Metabolite Structure Prediction Focused on the Cytochrome P450 Enzyme Family

Christina de Bruyn Kops[1]*, Conrad Stork[1], Nikolay Kochev[2,3], Nina Jeliazkova[2], Johannes Kirchmair[1,4,5]

[1] *Universität Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Department of Computer Science, Center for Bioinformatics, Hamburg, Germany*

[2] *Ideaconsult Ltd, Sofia, Bulgaria*

[3] *University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry, Plovdiv, Bulgaria*

[4] *University of Bergen, Department of Chemistry, Bergen, Norway*

[5] *University of Bergen, Computational Biology Unit, Department of Informatics, Bergen, Norway*

**kops@zbh.uni-hamburg.de*

Biotransformation of small organic molecules can result in metabolites with greatly modified biological and physicochemical properties compared to those of the parent compound; [1] hence, knowledge of the metabolic fate of xenobiotics in humans is vital for the development of safe and effective drugs. Regioselectivity prediction, the prediction of the locations in the molecule where metabolic reactions are initiated, is an aspect of metabolism prediction that is often an aim in and of itself but can also be applied as an initial step towards predicting the structures of the metabolites.

Here we examine the effect of an initial regioselectivity prediction step on the quantity and quality of metabolite structures that can be generated. We have developed a strategy for metabolite structure prediction that is based on FAME 2, [2] our recently-developed and highly effective machine learning method for human cytochrome P450 (CYP) regioselectivity prediction. As part of this strategy, we have assembled a thorough collection of known CYP-mediated reactions based on the scientific literature and chemical knowledge. By applying these reactions to the sites of metabolism predicted by FAME 2, we are able to correctly predict the majority of known metabolites while lowering false-positive prediction rates compared to the application of the reactions to all atom positions in the parent compounds.

[1] J. Kirchmair, A. H. Göller, D. Lang, J. Kunze, B. Testa, I. D. Wilson, R. C. Glen, G. Schneider, *Nature Rev Drug Discov* **2015**, 14, 387-404.

[2] M. Šícho, C. de Bruyn Kops, C. Stork, D. Svozil, J. Kirchmair, *J Chem Inf Model* **2017**, 57, 1832-1846.