

**Exploring chemical space beyond GDB17**J. Arús-Pous<sup>1</sup>, R. Visini<sup>1</sup>, M. Awale<sup>1</sup>, J. L. Reymond<sup>1\*</sup><sup>1</sup>University of Bern

The chemical space contains all possible molecules. We are interested in characterizing and studying this space to understand its properties and to aid in the process of drug discovery (such as finding new and unexplored chemotypes). In order to do that, we exhaustively sample parts of it. The GDB database family [1], which are a set of databases that contain all possible organic molecules with a given number of heavy atoms or less, has proven extremely successful and we have been able to describe and study most of the chemical space of drug-like molecules with 11 ( $\sim 10^7$ ) [2], 13 ( $\sim 10^9$ ) [3] and 17 ( $\sim 10^{11}$ ) [4] or less atoms. The huge size of both databases make the widely used chemoinformatics methods not feasible, so we are constantly developing new methods to work with such enormous databases [5]. Currently, we are working on the exploration of drug-like molecules that are bigger than 17 atoms [5]. As the size of such databases would easily exceed the available computational power (GDB-20 would be  $\sim 10^{18}$ ), we are sampling more specialized databases. Also, to be able to work with this large amount of molecules, we are currently developing machine learning technologies that will enable us to perform better searches on the databases.

[1] J-L Reymond, *Acc. Chem. Res.*, **2015**, 48, 722-730

[2] T. Fink, H. Bruggesser, J.-L. Reymond, *Angew. Chem. Int. Ed.*, **2005**, 44, 1504-1508

[3] L. C. Blum, J.-L. Reymond, *J. Am. Chem. Soc.*, **2009**, 131, 8732-8733

[4] L. Ruddigkeit, R. van Deursen, L. C. Blum and J.-L. Reymond, *J. Chem. Inf. Model.*, **2012**, 52, 2864-2875

[5] M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.*, **2014**, 54, 1892-1907