Comparative Analysis of the Chemical Space of Known and Purchasable Natural Products

Ya Chen [1], Marina Garcia de Lomana [1], Nils-Ole Friedrich [1], Johannes Kirchmair [1]

[1] *Universität Hamburg, MIN Faculty, Department of Informatics, Center for Bioinformatics, Hamburg, Germany*

Natural products cover a large chemical space and exhibit a wide range of bioactivities, making them an important resource for drug discovery [1-3]. We have recently analyzed the value of 25 virtual and 31 physical natural product libraries for computer-guided drug discovery [4]. These libraries cover a total of 250,000 natural products, at least 10% of which are readily purchasable from commercial sources.

In this contribution, we present a detailed analysis of the physicochemical property space of natural products that extends beyond the reach of earlier reports. We implemented a new algorithm ("SugarBuster") that identifies and removes sugars and sugar-like moieties, which are generally undesired in the context of drug discovery, from natural products. Use of this algorithm gives a more realistic view of the physicochemical properties of aglycons that may serve as templates for drug design. We also compare the physicochemical properties and scaffold diversity of purchasable natural products to those of all known natural products. This analysis provides valuable insights into the relevance of purchasable natural products for drug discovery and points out areas in the chemical space that are only covered by a subset of natural products requiring more demanding and expensive sourcing. Furthermore, we implemented a rule-based approach for the automated recognition of several structural classes of natural products (e.g. alkaloids or flavonoids), which allowed us to quantify their abundance among various data sources.

[1] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2016**, 79, 629–661.
[2] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nat. Rev. Drug Discov.* **2015**, 14, 111–129.
[3] T. Rodrigues, D. Reker, P. Schneider, G. Schneider, *Nat. Chem.* **2016**, 8, 531–541.
[4] Y. Chen, C. de Bruyn Kops, J. Kirchmair, *J. Chem. Inf. Model.* **2017**, 57, 2099–2111.