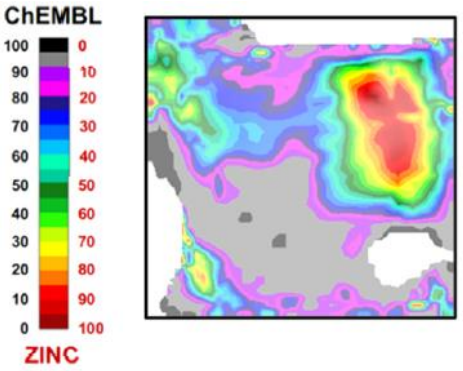
Chemography Concept in Chemical Space Exploration

Alexandre Varnek
Laboratory of Chemoinformatics, University of Strasbourg, France

Huge volume of biomedical data in chemistry and life sciences requires development of new methods and approaches for their analysis. Nowadays, some 200 million compounds are recorded in public and commercial databases and this is still a very small portion of chemical universe estimated as ~ 10^{33} drug-like molecules [1] which could potentially be synthesized. *Chemography* is a relatively new field dealing with visualization, analysis and modeling of chemical data. Its implementation in Generative Topographic Mapping (GTM) method is particularly well suited to treat big chemical data [3]. In GTM, compounds encoded by *N* molecular descriptors are represented on two-dimensional maps which can be efficiently used for chemical data visualization and analysis, prediction of physico-chemical properties or biological activities, comparison of large databases of chemical compounds, drugs repurpusing and virtual screening [2]. Being integrated with LSTM encoencoder, GTM can also be used for automatized generation of chemical structures with dezired properties [4] and new chemical transformations [5]. Some examples of GTM application to discovery of new molecules and reactions will be discussed.



This map accomodates >515 M structures of drug-like compounds from ZINC and ChEMBL databases [3]. One can see that compared to ChEMBL, large ZINC database occupies relatively small area in the chemical space.

References.

1. P.G. Polishchuk, T.I. Madzhidov, A. Varnek *J. Comput. Aided Mol. Des.*, 2013, **27**:675–679
2. D. Horvath, G. Marcou, A. Varnek, *Drug Discovery Today: Technology*, 2019, 32–33, 99-107
3. Y. Zabolotna, A. Lin, D. Horvath, G. Marcou, D. Volochnyuk, A. Varnek *J. Chem. Inf. Model.* 2021, 61, 1, 179–188
4. B Sattarov, I. Baskin, D. Horvath, G. Marcou, EJ Bjerrum, A. Varnek, *J. Chem. Inf. Model.*, 2019, **59**(3), 1182-1196.
5. W. Bort, I. Baskin, T. Gimadiev, A. Mukanov, R. Nugmanov, P. Sidorov, G. Marcou, D. Horvath, O. Klimchuk, T. Madzhidov and A. Varnek *Nature Scientific Reports*, 2021, DOI: 10.1038/s41598-021-81889-y