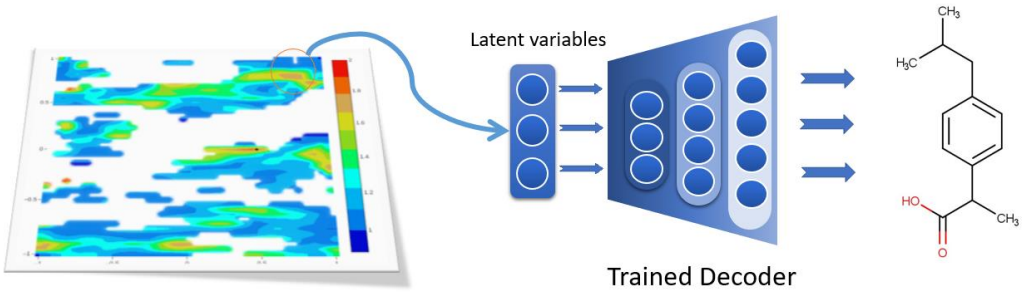
[P14] Deep Learning and Generative Topographic Mapping tandem in de novo design of biologically active compounds

Boris Sattarov[1], Dragos Horvath[1], Esben Jannik Bjerrum[2], Gilles Marcou[1], Igor Baskin[1,3] and Alexandre Varnek [1]

[1] *Laboratory of Chemoinformatics, UMR 7140, University of Strasbourg, France*
[2] *Wildcard Pharmaceutical Consulting, Søborg, Denmark.*
[3] *Faculty of Physics, Moscow State University, Russia*

Here we show that Generative Topographic Mapping (GTM) [1] can be used to explore the latent space of the SMILES-based autoencoders and generate on target-focused molecular libraries. We have built a sequence-to-sequence neural network with bidirectional Long Short-Term Memory [2,3] layers and trained it on SMILES strings from ChEMBL23. This method converts SMILES into a latent vector of 128 real values. This vector "encodes" all the structure information, and therefore can be "decoded" back to the original SMILES. Very high reconstruction rates of test set molecules were achieved (>98%). The key benefit is that "navigating" the chemical space (the enumeration of possible compounds, a difficult problem) can be reduced to navigating the latent space (easily feasible *in Silico*) followed by "stepping back" to the associated SMILES strings of likely very original structures.

Using GTM, we have projected the latent space on a 2D map, in which key zones, enriched in "active" compounds can be highlighted. The chemical space area preferentially populated by A2A (adenosine receptor) inhibitors was targeted by means of an evolutionary process in the latent space. The visited latent space points were projected on the GTM, to verify whether they match the A2A target zone. If likelihood to be active is high, the latent space point is verified to match a chemically valid structure (by both RDKit's synthetic feasibility score, combined with ChemAxon's Structure Checker diagnosis). *Ab initio* generated compounds satisfying these criteria were confronted to A2A pharmacophore models and eventually docked into the receptor site. In parallel, their originality was assessed by a search of nearest neighbors in the PubChem database. Fully original structures were shown to successfully dock into the target, pending experimental validation.



Bibliography:

[1] Kireeva et al. (2012) Molecular informatics, 31(3-4), 301-312.
[2] Gómez-Bombarelli et al (2016). ACS Central Science.
[3] Xu et al. (2017,). Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 285-294). ACM.