P.Sidorov[1,2]
H.Gaspar[1]
G.Marcou[1]
D.Horvath[1]
A.Varnek[1]

Mappability of Drug-like Space: towards a polypharmacologically competent map of drug-relevant compounds

[1] Laboratoire de Chémoinformatique, UMR 7140 CNRS - Université de Strasbourg, 1 rue Blaise Pascal, Strasbourg 67000, France;
[2] Laboratory of Chemoinfomatics, Butlerov Institute of Chemistry, Kazan Federal University, Kazan, Russia
pavel.sidorov@unistra.fr

This work attempts to address the question whether a “Universal model” of the Chemical Space exists and propose a representation of it. A universal model is intended as a probability distribution of compounds that could be set-independent. The probabilistic model is build as a Generative Topographic Map (GTM). The claim of “universality” is quantitatively justified, with respect to all the structure-activity information available so far. To this purpose, an evolutionary map growth and selection procedure considered both the choice of meta-parameters (poling molecule sets, descriptor types) and map-specific parameters (size, RBF function controls, etc) as degrees of freedom. It was associated to a fitness function measuring the polypharmacological performance of the map, with respect to a multi(144)-target quantitative affinity prediction challenge. Under the pressure of Darwinian selection, the emerging maps were pushed to find (a) the best descriptor type, out of the proposed substructural molecular fragments descriptors schemes, and (b) the specific non-linear “recipe” of generating a model GTM probability distribution which enhances the information contained in certain descriptor elements, but suppresses descriptor “noise”. The fittest manifolds were seen to “grow” in rather low-resolution molecular descriptor spaces: pharmacophore- or force-field-type colored atom pairs and triplets rather than more specific sequence or circular fragment counts included in the pool of competing ISIDA descriptor types.

Obtained maps were perfectly suited to solve classification problems: on the overall, more than 80% of the more than 600 distinct and varied classification problems, chosen such as to cover a maximum of exploitable SAR data, were successfully solved. This justifies, in our view, the claim of "Universality" of the constructed GTMs.

In addition, intuitive 2D representations were shown to provide an insightful analysis of drug-like space, and provide huge perspectives for target- and therapeutic range-related compound collections. Due to quantitative validation, the user may gain confidence in the rendered visual patterns, and draw very meaningful conclusions on their behalf.

P.Sidorov thanks the Program of Competitive Growth of Kazan Federal University for the support