Journal of Computer-Aided Molecular Design 2015 vol.29 N12, pages 1087-1108

Mappability of drug-like space: Towards a polypharmacologically competent map of drug-relevant compounds

Sidorov P., Gaspar H., Marcou G., Varnek A., Horvath D.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2015 Springer International Publishing Switzerland. Intuitive, visual rendering - mapping - of high-dimensional chemical spaces (CS), is an important topic in chemoinformatics. Such maps were so far dedicated to specific compound collections - either limited series of known activities, or large, even exhaustive enumerations of molecules, but without associated property data. Typically, they were challenged to answer some classification problem with respect to those same molecules, admired for their aesthetical virtues and then forgotten - because they were set-specific constructs. This work wishes to address the question whether a general, compound set-independent map can be generated, and the claim of "universality" quantitatively justified, with respect to all the structure-activity information available so far - or, more realistically, an exploitable but significant fraction thereof. The "universal" CS map is expected to project molecules from the initial CS into a lower-dimensional space that is neighborhood behavior-compliant with respect to a large panel of ligand properties. Such map should be able to discriminate actives from inactives, or even support quantitative neighborhood-based, parameter-free property prediction (regression) models, for a wide panel of targets and target families. It should be polypharmacologically competent, without requiring any target-specific parameter fitting. This work describes an evolutionary growth procedure of such maps, based on generative topographic mapping, followed by the validation of their polypharmacological competence. Validation was achieved with respect to a maximum of exploitable structure-activity information, covering all of Homo sapiens proteins of the ChEMBL database, antiparasitic and antiviral data, etc. Five evolved maps satisfactorily solved hundreds of activity-based ligand classification challenges for targets, and even in vivo properties independent from training data. They also stood chemogenomics-related challenges, as cumulated responsibility vectors obtained by mapping of target-specific ligand collections were shown to represent validated target descriptors, complying with currently accepted target classification in biology. Therefore, they represent, in our opinion, a robust and well documented answer to the key question "What is a good CS map?

http://dx.doi.org/10.1007/s10822-015-9882-z

Keywords

Chemical space mapping, Generative topographic maps, Polypharmacology, Structure-property relationships