Pharmacological Space Mapping via Structure-Target Relationship using Layered SOMs

Nicoleta STOIAN[a,b*], Liliana HALIP[a], and Laura PITULICE[b]

[a] Coriolan Dragulescu Insititute of Chemistry, Bvld. Mihai Viteazu no. 24, 300223 Timişoara, Romania.
[b] Chemistry Department, Faculty of Chemistry, Biology, Geography, West University of Timişoara, Bvld. Vasile Parvan no. 4, 300223 Timişoara, Romania.
E-mails: (*) nicoleta.stoian01@e-uvt.ro; lili.halip@gmail.com; laura.pitulice@e-uvt.ro

* Author to whom correspondence should be addressed;

Abstract

Conventional methods often fail to solve the divergence between chemical structure and complex biological response, being a significant challenge in computational pharmacology. Consequently, there is a growing need for integrative models that surpass the limitations of traditional structure-activity analysis through multidimensional data fusion. Based on the premise that chemical and biological similarities capture distinct but complementary dimensions of drug mechanisms of action, we developed a dual-layer Self-Organizing Map (SOM) model. This architecture combines two complementary representations for 1,600 approved drugs: a binary chemical fingerprint that encodes each compound's structure, and a continuous, biologically informed fingerprint derived through deep learning from target protein sequences. This hierarchical approach preserves the domain-specific topology of each space before integrating them into a unified map and facilitated the identification of functional convergences, where structurally heterogeneous compounds are projected into identical regions of biological activity. Model validation was performed using rigorous internal metrics, including quantization error and topographic error. By simultaneously projecting chemical structure and known target activity profiles, this workflow provides a robust tool for in silico pharmacological profiling, enabling the detection of drug repurposing opportunities or target fishing of early-stage drug candidates. Clustering analysis revealed five distinct groups of compounds differentiated primarily by molecular size, lipophilicity, polarity, and target class distribution. Clusters show well-defined structural motifs, larger and more polar enzyme-associated compounds being separated from smaller, less polar GPCR-oriented molecules, as well of a distinct group of more lipophilic nuclear receptor ligands. Differences in physicochemical properties were accompanied by variations in target promiscuity, with some clusters exhibiting broader polypharmacology and others showing greater selectivity.

Keywords: Self-organizing maps; MACCS fingerprint; ProtBERT protein embeddings; Drug repurposing.