CoLiNN: Enabling Rapid Visualization of DNA-Encoded Libraries Without Compound Enumeration

Regina Pikalyova, Tagir Akhmetshin, Dragos Horvath, Alexandre Varnek

Laboratory of Chemoinformatics, University of Strasbourg, 67000, Strasbourg, France.

The analysis of vast combinatorial libraries requires efficient chemical space visualization tools. The traditional workflow involves the enumeration of compounds, their standardization, calculation of molecular descriptors, and application of dimensionality reduction techniques, all of which are computationally intensive. To streamline this workflow, we developed the Combinatorial Library Neural Network (CoLiNN), a graph convolutional network. CoLiNN predicts positions of molecules on a 2D map directly from the corresponding building blocks and reactions, without the need for compound enumeration and descriptor computation (see figure 1).

In this work, CoLiNN was used to predict Generative Topographic Maps (GTMs) for 2,5K DNA-Encoded Libraries[1] (DELs) with diverse reaction schemes. It demonstrated high accuracy in predicting GTMs for libraries not included in the training set, significantly reducing computational time and storage requirements. Thus, CoLiNN provides a rapid and scalable approach for the visualization and analysis of ultra-large combinatorial libraries.

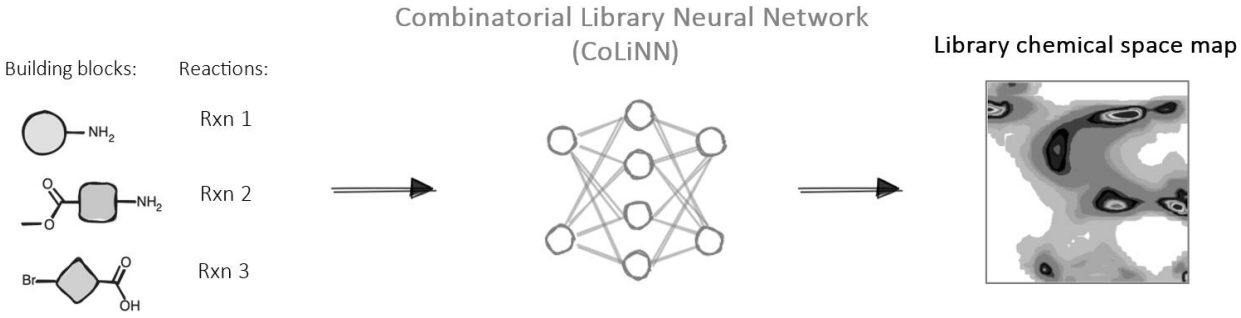

Figure 1. Scheme illustrating the main idea of CoLiNN: Prediction of the chemical space map from only building blocks and reactions of a combinatorial library without compound enumeration.

Bibliography :

[1] Lerner, R. A.; Brenner, S. DNA-Encoded Compound Libraries as Open Source: A Powerful Pathway to New Drugs. Angewandte Chemie International Edition 2017, 56 (5), 1164–1165.