Gene expression data-driven scaffold-constrained molecular structure generation by deep neural network

Yuki Matsukiyo, Kyushu Institute of Technology

In the lead optimization process, modifying functional groups while retaining a scaffold (i.e., core structure) of a molecule is one of the promising ways to enhance molecular properties. To accelerate this process, many scaffold-constrained molecular generation algorithms using deep neural networks such as recurrent neural network, variational autoencoder, and generative adversarial network have been proposed, but most previous scaffold-constrained molecular generation algorithms focused on generating chemically valid molecules. As an alternative approach, the use of biological big data such as gene expression profiles has been recently proposed to generate molecules. However, there is no study on scaffold-constrained molecular generation using gene expression profiles. In this study, we propose a novel computational method to generate molecules as candidates for regulators of a therapeutic target protein from target-perturbed gene expression profiles in a scaffold-constrained manner. We used a target-perturbed gene expression profile and a molecular scaffold as inputs, and generated new inhibitor or activator candidate molecules for each therapeutic target protein. Furthermore, we evaluated the validity of generated molecules using docking simulations on the therapeutic target protein structure. Our proposed method is expected to contribute to more efficient lead optimization.