Inverse molecular design via graph-based conditional variational autoencoder

Hyunseung Kim [1] and Won Bo Lee[1*]

[1] *School of Chemical and Biological Engineering, Seoul National University, Seoul 08826, Republic of Korea*

Corresponding author (Electronic mail: wblee@snu.ac.kr)

Inverse molecular design to discover a molecular structure that satisfies desired properties is one of the challenging problems in material discovery. Traditionally, experts have searched for chemical spaces based on expert knowledge and intuition, however, this is not efficient for quickly exploring vast spaces. For this reason, methodologies have been proposed that enable effective exploration of vast spaces by utilizing machine learning which is specialized in data pattern recognition, mapping, and inference. In this study, a model that uses a graph-based conditional variational autoencoder to design the molecular structure that satisfies the desired properties (logarithm of partition coefficient, topological polar surface area, and quantitative estimation of drug-likeness) is proposed. The proposed model is trained and tested using ZINC 15, a small molecule drug database.

References

[1] A. M. Virshup et al., "Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds," *J. Am. Chem. Soc.*, **135**, 7296 (2013)

[2] S. Ryu et al., "Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network," *arXiv preprint*, arXiv:1805.10988 (2018)