Deep learning-based molecular design with high synthetic feasibility

Woo Youn Kim*

Department of Chemistry and AI Institute, KAIST, Republic of Korea
*e-mail address: wooyoun@kaist.ac.kr

Deep generative models are attracting great attention for molecular design with desired properties. Most existing models generate molecules by sequentially adding atoms. This often renders generated molecules with less correlation with target properties and low synthetic accessibility. Here, we propose a building block-based molecular generative model which designs new molecules with target properties by sequentially adding molecular building block to any given starting molecule. A key feature of our model is a high generalization ability in terms of property control and building block types. The former becomes possible by learning the contribution of individual fragments to the target properties in an auto-regressive manner. For the latter, we used a deep neural network that predicts the bonding probability of two molecules from the embedding vectors of the two molecules as input. The high synthetic accessibility of the generated molecules is implicitly considered while preparing the building block library with the BRICS decomposition method. We show that the model can generate molecules with the simultaneous control of multiple target properties at a high success rate. As a practical application, we demonstrated that the model can generate potential inhibitors with high binding affinities against the 3CL protease of SARS-COV-2 in terms of docking score.