Title

A Target Protein-focused 3D Molecular Generative Framework Toward Generalized Structure-based Drug Design

Authors

Wonho Zhung,[1] Hyeongwoo Kim,[1] Joongwon Lee,[1] **Woo Youn Kim***[1,2,3]

Affiliation

[1]Department of Chemistry, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[2]HITS Incorporation, 124 Teheran-ro, Gangnam-gu, Seoul 06234, Republic of Korea
[3]KI for Artificial Intelligence, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

Submission Type

Oral presentation

Presenting Area

Machine Learning

Abstract

Generative models for *de novo* molecular design have been the subject of immense interest in drug discovery, reducing the time and cost to identify hits from the vast chemical possibilities. Low generalizability, however, is still a big huddle to develop a generative model for structure-based drug design where the data is often sparse and biased, leading to poor-quality samples designed for the unseen target protein. Here, we propose a target protein-focused 3D molecular generative framework that explicitly models structure-activity relationships (SARs) between the protein and the ligand to improve the model generalization. The framework adopts the encoder-decoder architecture, where the encoder maps the local 3D context of binding pockets to a latent space while the decoder sequentially grows the ligand from the latent vector to fulfill the given SAR conditions. We show that this strategy enables our framework to achieve generalizability and interpretability by unveiling the model's decision in each sampling step and evaluating the contribution of generated protein-ligand interactions in binding stability via short MD simulation. Furthermore, we demonstrate the effect of using different SAR conditions on the sample distribution in terms of binding affinity and enrichment factor toward the target protein to control the potency and selectivity of designed ligands.