[BIOINFO 2025]

Aligning Molecules and Fragments in a Shared Embedding Space for RL-Based Molecule Generation

Youngkuk Kim[1], Yinhua Piao[2], Sangseon Lee[3], and Sun Kim[1,4,5,6,*]

[1]*Department of Computer Science and Engineering, Seoul National University*
[2]*Bio-MAX Institute, Seoul National University*
[3]*Department of Artificial Intelligence, Inha University*
[4]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[5]*Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[6]*AIGENDRUG Co., Ltd.*
**Corresponding author: sunkim.bioinfo@snu.ac.kr*

Drug discovery is a complex and resource-intensive process requiring the design of molecules that possess specific chemical and biological properties, such as high binding affinity and drug-likeness. Fragment-based drug discovery (FBDD) has gained prominence as a strategy for efficiently identifying lead compounds by deconstructing molecules into smaller fragments. However, existing approaches face challenges in fully leveraging the relationships between molecules and their constituent fragments, especially in optimizing molecular properties. In this paper, we introduce Molecule-Fragment Representation Alignment space for RL-based Generation (M-FRAG), a novel framework that harmonizes molecule and fragment embeddings in a shared, property-driven space. By aligning fragments with their molecular context, M-FRAG ensures that fragment selection is optimized both for chemical feasibility and the desired molecular properties. Using reinforcement learning, M-FRAG generates chemically realistic molecules optimized for target properties while also providing interpretability for individual fragments during the molecule generation process. Experimental results demonstrate that M-FRAG outperforms existing methods in terms of optimization, diversity, and chemical validity, positioning it as a powerful tool for the efficient and transparent generation of drug-like molecules.