**[BIOINFO 2024]**

# MV-CLAM: Multi-View Molecular Interpretation with Cross-Modal Projection via Language Model

Sumin Ha[1,†], Jun Hyeong Kim[2,†], Yinhua Piao[3], and Sun Kim[1,2,3,4]

[1] *Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[2] *Bio-MAX/N-Bio, Seoul National University*
[3] *Department of Computer Science and Engineering, Seoul National University*
[4] *AIGENDRUG Co., Ltd.*
*† These authors contributed equally to the work*
** Corresponding author: sunkim.bioinfo@snu.ac.kr*

Large language models (LLMs) have demonstrated significant promise in the biomolecular field, particularly in enhancing the quality of molecular captions. While effective adaptations of enriched molecular representations have driven cross-modal research, previous studies have largely focused on aligning a single molecular view with text. Naïve approaches to capture the richness of diverse modalities and dimensions require independent alignment of the multiple views, leading to separately aligned molecule and text representations. To tackle this issue and reduce computational workload, we introduce MQ-Former, a multi-querying transformer integrated within the LLM framework. This architecture features a novel cross-model projector that enables the simultaneous alignment of 2D and 3D molecular representations to a unified text token. By utilizing a shared self-attention layer, MQ-Former retains rich molecular embeddings across various dimensions while consolidating them into a single universal molecular token. The universal molecular token serves as a soft prompt for LLAMA2 with 1D SMILES string and instruction prompt. Our molecule domain specialized language model, MV-CLAM, surpasses baseline models in both molecule-text retrieval and molecular captioning tasks. Additionally, we explore further molecular interpretation through experiments in zero-shot molecule editing and molecule-related question answering. By effectively integrating multi-view molecular data into a format suitable for LLMs, our method enhances the characterization and understanding of chemical structures, facilitating a smoother transition from molecular data to textual descriptions.