Towards Interpretable Prediction of SGLT2 Inhibitor Activity via Flexibility- and Electronic-Structure-Enhanced Graph Neural Networks

Yusuke Tateishi
Kumamoto University

Accurate prediction of molecular activity based solely on static chemical structures remains challenging. To address this issue, we propose a graph neural network framework that augments molecular graphs by embedding node features derived from the ligand's flexibility and quantum chemical properties. Flexibilities, such as conformational diversity and vibrational displacements, provide insight into entropic effects. Meanwhile, electronic-structure features, including charge redistribution in different states (e.g., S0 to cation and S0 with the effect of water), capture reactivity and interaction capability with the target molecules. These additions not only aim to improve predictive performance but also to enhance interpretability by linking activity to chemically and physically meaningful molecular phenomena. As a case study, we applied our approach to the activity prediction of sodium–glucose cotransporter 2 (SGLT2) inhibitors, a class of ligands that interact with a flexible membrane protein, making the prediction particularly complicated. Our model predicted IC50 values of SGLT2 inhibitors with moderate accuracy. Additionally, this approach provides atom-level contributions, important subgraphs (i.e., fragments), and gradient-based feature importance per atom to the inhibitory prediction. This study contributes to interpretable machine learning grounded in chemical and physical perspectives in drug discovery by integrating fundamental structural information, conformational diversity, and electronic-structure properties.