

Modeling the Stereoelectronic Effects of Glycosylation Dynamics with Machine-Learned Interatomic Potentials

Eugene Chung¹, Mateusz Mariani^{1,2}

1. Ph.D Program in Chemistry, CUNY Graduate Center
2. Department of Chemistry, CUNY Hunter College

Glycosylation forms glycosidic bonds at the anomeric center of carbohydrates and is essential to biological processes and the synthesis of therapeutics and biomaterials. However, understanding and predicting the pathways and stereochemical outcomes of chemical glycosylation remain challenging. Solvent, donor structure and conformation, protecting groups, and activation conditions can favor competing reaction pathways and influence product selectivity. Ab initio molecular dynamics (AIMD) and metadynamics in explicit solvent can provide atomistic insight into these effects, but their computational cost limits the extent of sampling.

Here, we develop a machine-learned interatomic potential (MLIP) to reduce the computational cost of simulating chemical glycosylation. We use approximately 200 ps of AIMD trajectories generated with CP2K to construct training, validation, and test datasets containing atomic coordinates, energies, forces, and partial charges. Starting from the AIMNet 2025 foundation model, we fine-tune the potential for chemical environments relevant to glycosylation.

We evaluate the fine-tuned model by comparing MLIP-driven trajectories with reference AIMD trajectories generated using CP2K, focusing on the behavior of the leaving group at the anomeric carbon and the protecting groups on the sugar ring. The model successfully reproduces the behavior of both observed in the reference simulations, providing initial evidence that it captures local molecular dynamics relevant to glycosylation. These results support further development of MLIP-driven simulations for more extensive sampling, with the longer-term goal of investigating how molecular structure and reaction conditions influence reaction pathways and stereoselectivity.