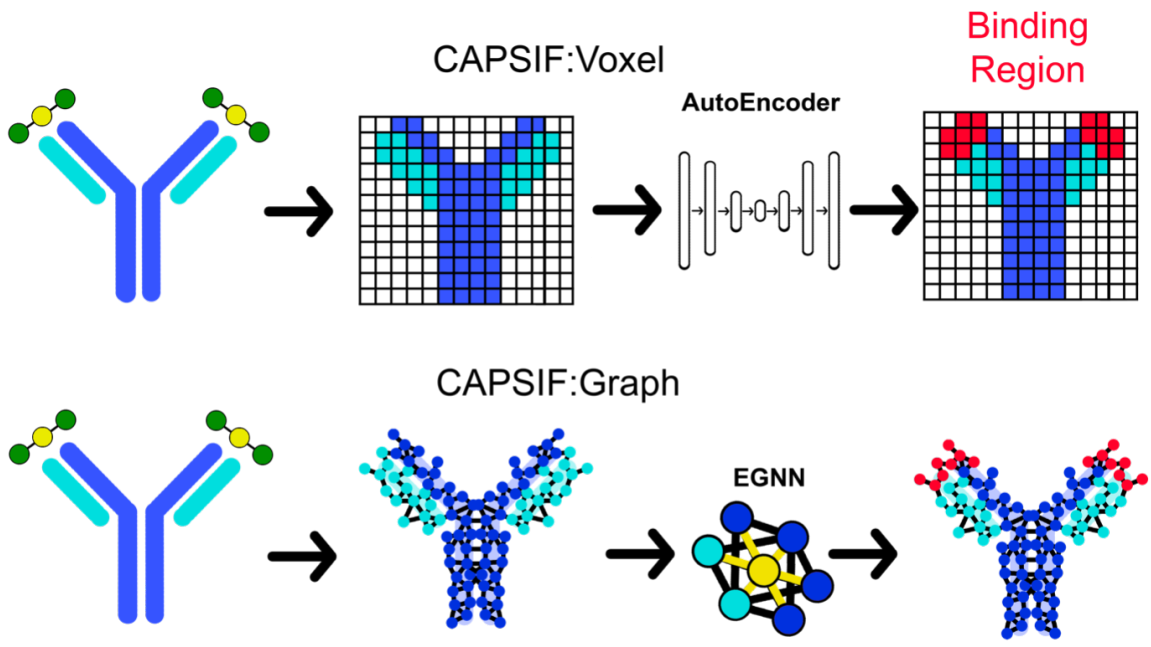
Structure based neural network predictions of protein carbohydrate interactions

Samuel CANNER [1], Sudhanshu SHANKER [1], Jeffrey J. GRAY [1]

[1] The Johns Hopkins University, Baltimore, MD 21218, [2] Department of Chemistry, UBC Faculty of Science, 2036 Main Mall, Vancouver, BC, Canada

scanner1@jhu.edu

Carbohydrates dynamically and transiently interact with proteins for cell-cell recognition, cellular differentiation, immune response, and other cellular processes. Despite the molecular importance of these interactions, there are currently few reliable computational tools to predict potential carbohydrate binding sites on any given protein. Here, we present two deep learning (DL) models named CArbohydrate-Protein interaction Site IdentiFier (CAPSIF) that predicts carbohydrate binding sites on proteins: (1) a 3D-autoencoder voxel-based neural network model (CAPSIF:V) and (2) an equivariant graph neural network model (CAPSIF:G). We found CAPSIF:V performs better than CAPSIF:G, and both models outperform previous surrogate methods used for carbohydrate binding site prediction. CAPSIF also performs well when starting from computationally determined structures from AlphaFold2. CAPSIF models can be used in conjunction with local glycan-docking protocols, such as GlycanDock, to predict bound protein-carbohydrate structures. We will share how these tools can be used for *ab initio* predictions of bound oligosaccharide transferase-carbohydrate structures.



CArbohydrate Protein Site IdentiFier (CAPSIF) neural network architectures