

Protein structure prediction algorithms such as AlphaFold2 and ESMFold have dramatically increased the availability of high-quality models of protein structures. Because these algorithms predict only the structure of the protein itself, there is a growing need for methods that can rapidly screen protein structures for ligands. In this work, we introduce SE3Lig, a model for semantic in-painting of small molecules in protein structures. Specifically, we report SE(3)-equivariant CNNs trained to predict the atomic densities of common classes of cofactors (hemes, flavins, etc.) and the water molecules and inorganic ions in their vicinity. While the models are trained on high-resolution crystal structures of enzymes, they perform well on structures predicted by AlphaFold2, which suggests that the algorithm correctly represents cofactor-binding cavities.