Model Report

Description of the Model

Type of model and method

Our submission uses the publicly released ESMFold2 protein and ligand co-folding model in inference mode. We did not train or fine-tune any model.

For each of the 184 test ligands we co-folded the PXR ligand-binding domain, residues 142 to 434 of UniProt O75469, with the ligand SMILES string as provided by the challenge. The protein input includes a multiple sequence alignment of approximately 2,500 unique sequences assembled from public databases with the ColabFold search server and truncated to a depth of 1,024. Each sampling run uses ten recycling loops and draws five diffusion samples over 68 denoising steps. Repeating this across many random seeds accumulates a pool of approximately 2,300 candidate poses per ligand.

From each ligand's pool we selected the pose with the highest iPTM score, which is ESMFold2's own estimate of the accuracy of the predicted protein and ligand interface. A conservative consensus correction was then applied. For each ligand we compared this main pick against four alternative sources of a single best pose. Three were our own alternative sampling configurations, namely alignment depth 256, twenty recycling loops, and an earlier configuration that omits the alignment. The fourth was a publicly released set of predictions shared by another challenge participant, generated with the same co-folding model without an alignment. Where the main configuration's pose was geometrically isolated, while at least two of the alternative sources agreed with each other within 1.5 Å ligand RMSD, we replaced the main pick with a pose from the agreeing cluster. This rule replaced the pose for six of the 184 ligands, five with the pose from the publicly released set and one with the pose from our depth-256 configuration.

Every candidate pose was screened by rebuilding the ligand's bond graph from its atomic coordinates with OpenStructure and requiring an exact element-labeled graph match to the provided SMILES, together with a minimum heavy-atom separation of 1.0 Å. When the most confident pose failed this screen, the next most confident passing pose was selected instead.

Finally, all 184 assembled models were audited for structural defects with the structure-preparation step of OpenStructure, which rebuilds each ligand's bond graph from its atomic coordinates, and with a measurement of the distance from each ligand's centroid to the binding pocket. The audit found six defective models. In four of them, no sampled pose had passed the bond-graph screen, because atoms sat close enough that the rebuilt graph gained one or two bonds and no longer matched the provided SMILES, which prevents automatic ligand atom assignment during scoring. In three of them, one overlapping with the first group, the ligand lay 35 to 49 Å outside the binding pocket. We replaced these six poses with

structurally sound alternatives. Two came from Boltz-2 co-folding guided by six public PXR template structures from the Protein Data Bank, selected by Boltz-2's own confidence score. Two came from additional ESMFold2 sampling, selected by iPTM. The final two were chosen from the ESMFold2 pool by taking, among the twenty most confident poses, the pose with the most neighbors within 1.5 Å ligand RMSD. After these repairs, every ligand's rebuilt bond graph exactly matches its SMILES and every ligand centroid lies within 5.2 Å of the pocket center.

The output for each ligand is one PDB file containing the protein chain and the ligand as a HETATM residue named LIG, rigid-body aligned on 38 binding-site alpha carbons to a common PXR reference frame. All 184 outputs were validated with the tutorial structure checker before submission.

Pre-training and external data

All tools named above are publicly available and were used without modification. No proprietary data or models were used. The challenge inputs to the submission are the test-set SMILES strings and the protein sequence. The six template structures used for the two Boltz-2 replacement poses, and the local benchmark described below, are public crystal structures from the Protein Data Bank.

Hyperparameter optimization

The decision to include a multiple sequence alignment was made on the local benchmark described below, where it improved both pose quality and the reliability of the confidence score. The alignment depth, the number of recycling loops, and the size of the pose pool were then compared directly on the public leaderboard across five submissions of configuration variants, and the best was retained. The consensus correction was designed to fire rarely, and an analogous criterion on the local benchmark selected two of 59 structures. The repair step has no tuned parameters, because it replaces only models with objectively verifiable defects. No systematic hyperparameter search was performed.

Molecular descriptors and data augmentation

No classical molecular descriptors are computed, and no protonation or tautomer enumeration is performed. The ligand input is the SMILES string exactly as provided. Repeated diffusion sampling across random seeds serves as the only augmentation, and the model's own confidence score chooses among the resulting poses.

Performance comments

For local method screening we assembled a benchmark of approximately 60 public PXR holo crystal structures from the Protein Data Bank and scored predictions against re-refined references using the LDDT-PLI implementation in OpenStructure, which is the challenge metric. This benchmark was useful for rejecting clearly inferior generators but was a poor guide to test-set performance, and gains measured on it repeatedly failed to transfer. We attribute this to

the benchmark structures falling within the training distribution of the co-folding models while most test ligands are chemically novel. In particular, we found that agreement between independently configured predictors on benchmark structures deposited before the models' training cutoffs largely reflects memorization of those structures rather than independent evidence of a correct pose. Consistent with this, the consensus correction described above is applied only in a deliberately restrictive form that modifies six of the 184 ligands.