**New tractability columns description**

Kristy Carpenter

May 14, 2025

has\_pdb\_structure

Input: Ensembl id

Output: 1 if there is an experimentally-determined structure for this target, otherwise None

Source: [Protein Data Bank](https://www.rcsb.org/) ([ref](https://doi.org/10.1093/nar/28.1.235))

Description: Identifies whether or not there is a structure for this target in the Protein Data Bank.

Why this is useful: We have more confidence in an experimentally-determined structure than a computationally-predicted one, and even if the structure does not have a ligand interaction, we can use a structure for pocket finding, docking, and structure-based drug development.

Relevant existing OTP annotations: “Structure with Ligand” is the closest existing OTP tractability annotation, but this does not capture experimental structures without a ligand, which are still very valuable and are not yet present for all targets.

has\_af2\_structure

Input: Ensembl id

Output: The average per-residue pLDDT (confidence) score for the highest-confidence AlphaFold2-predicted structure of this target, or None if there is no predicted structure

Source: [AlphaFold Protein Structure Database](https://alphafold.ebi.ac.uk/) ([ref](https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab1061/6430488))

Description: Identifies whether or not there is a computationally-predicted structure for this target in the AlphaFold Protein Structure Database. Structures are predicted by AlphaFold2. The per-residue confidence metric originally scales from 0 to 100, but here we re-scale it from 0 to 1. Per the AlphaFold Protein Structure Database website, some general guidelines for interpreting the confidence metric (rescaled from 0 to 1 here) are: greater than 0.9 is very high confidence, between .7 and .9 is high confidence, between 0.5 and 0.7 is low confidence, and below 0.5 is very low confidence.

Why this is useful: If there is not an experimentally-determined structure of a protein, then a high-confidence computationally-predicted structure can still be useful for identifying potential binding pockets. Knowing the overall confidence of the predicted structure helps us understand how much we can trust it for downstream applications.

Relevant existing OTP annotations: “Structure with Ligand” is the closest OTP tractability annotation, but this only captures experimentally-determined structures (with a ligand). In the post-AlphaFold era, we can obtain high-confidence structures computationally and use these to assess ligandability/druggability of a target. However, the confidence of predicted structures varies – not all predicted structures are equally good.

has\_biolip\_interaction

Input: Ensembl id

Output: 1 if there is a biologically-relevant ligand interaction for this target as annotated in the BioLiP database, in addition to a list of all biologically-relevant ligands and a brief description of each ligand; otherwise None

Source: [BioLiP](https://zhanggroup.org/BioLiP/) ([ref](https://academic.oup.com/nar/article/52/D1/D404/7233921))

Description: Identifies whether or not there is a BioLiP annotation for this target, and if so, what kinds of ligands are known to bind. Having a BioLiP annotation means that there is a structure in the Protein Data Bank for this target that has a biologically-relevant ligand interaction.

Why this is useful: Knowing a validated ligand is helpful both for identifying the binding site and what kind of ligand is able to bind. BioLiP differentiates biologically-relevant ligands from things like solvent interactions that would not yield useful information for druggability.

Relevant existing OTP annotations: “Structure with Ligand” is the closest OTP tractability annotation, and this column has a lot of overlap with BioLiP annotations. However, “Structure with Ligand” only captures if the target has a structure in which there is a small molecule ligand. The BioLiP annotations indicate whether the ligand is interacting with the given protein. For example, say there is a structure of the complex of Protein A, Protein B, and Ligand C, and in this structure, Ligand C is bound to Protein A. In OTP, the “Structure with Ligand” column will indicate an interaction for both Protein A and Protein B. The has\_biolip\_interaction column will only indicate an interaction for Protein A, not Protein B. Additionally has\_biolip\_interaction provides annotations on whether the ligands are DNA, RNA, peptide, organic small molecule, or inorganic small molecule. Lastly, “Structure with Ligand” flags non-biologically-relevant ligand interactions such as solvent interactions, and has\_biolip\_interaction does not.

has\_prank\_pdb\_pocket

Input: Ensembl id

Output: The highest probability score out of the set of all P2Rank-predicted pockets, or None if there are no predicted pockets

Source: [PrankWeb](https://prankweb.cz/) ([ref](https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0285-8))

Description: Identifies whether there are P2Rank pocket predictions for any PDB structure for this target (note that P2Rank predictions are not available over API for computationally-predicted pockets) and provides the probability score (from 0 to 1) of the highest-scoring pocket. These are computationally-predicted pockets based on a geometry-based pocket-finding strategy, scored by a random forest algorithm.

Why this is useful: P2Rank is an established pocket-finding method with good performance. It can be run on any protein structure (though it is easiest to access structures from the Protein Data Bank with the API, which is how we get scores in this script). Even if no ligand interaction has been observed through experimental structures, identifying a likely binding pocket on the surface of the protein is a useful starting point for assessing druggability.

Relevant existing OTP annotations: “High-Quality Pocket” and “Med-Quality Pocket” are the closest OTP tractability annotations. OTP currently uses the pocket finding method DrugEBIlity. While there are no evaluations that directly compare P2Rank and DrugEBIlity, adding another pocket prediction method may add value. Additionally, the has\_prank\_pdb\_pocket column returns a probability-based pocket score instead of a categorical score of “High” or “Med” which may aid comparison. The P2Rank algorithm is described in a published manuscript and can be used for any protein structure, whereas there does not appear to be documentation for DrugEBIlity publicly available online.