

PREDICTION EXAMPLES

In this document we will compare our predicted binding sites for three different proteins with the known structure of the binding site. All PDB files were obtained from the dataset of the publication by Ahmed et al. (2021). All example PDB files and prediction results can be found in the github: <https://github.com/aelhammad/deep-pocket.git>.

Example 1: Pentaerythritol Tetranitrate Reductase in complex with progesterone (PDB code: 1h60)

Pentaerythritol tetranitrate reductase degrades high explosive molecules including nitrate esters, nitroaromatics and cyclic triazine compounds. This enzyme also binds a variety of cyclic enones, including steroids; some steroids act as substrates whilst others are inhibitors (Barna et al., 2001). The next image was obtained after superimposing the structures of the protein and its known binding site, highlighted in red:



Figure 1. Example 1 with real known binding site

Our predicted binding site includes 11 residues, compared to the 99 residues of the known binding site. Out of these 11 residues, 3 are in common between the predicted and known binding sites. In the following image we can see the residues of our prediction highlighted in blue:

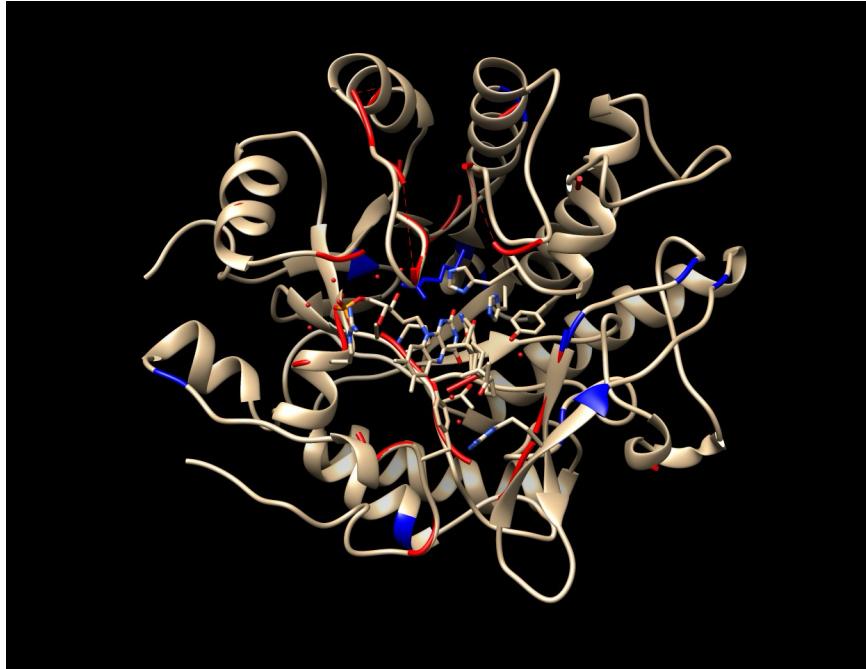


Figure 2. Example 1 with predicted binding sites

As we can see, even though the majority our predicted residues in the binding site are not the same as the ones in the known binding site, most are situated spatially close to residues that belong to the known binding site.

Moreover, we can see that a lot of the residues from the predicted and known binding sites are located in loop regions of the protein, which makes sense given the fact that the residues situated in loop regions have a bigger degree of conformational flexibility and surface accessibility compared to residues in secondary structure elements such as alpha helices or beta strands. This flexible nature of the loop regions makes them more prone to participate in the binding of other molecules. This also applies to the rest of the examples presented in this document.

Example 2: methotrexate-resistant Leu22Arg variant of mouse dihydrofolate reductase (PDB code: 1u72)

Dihydrofolate reductase catalyzes the reduction of folic acid to dihydrofolic acid and to tetrahydrofolic acid, an essential cofactor in the biosynthesis of thymidylate, purines and glycine. The folic acid analogue methotrexate has been shown to bind tightly to the active site of the enzyme, resulting in the death of exposed cells (Cody et al., 2005). The next image was obtained after superimposing the structures of the protein and its known binding site, highlighted in red:



Figure 3. Example 2 with real binding sites

Our model predicted only 3 residues compared to the 186 residues of the known binding site, without having any in common but being closely located to residues in the known binding site.

In these cases, we can lower the trust level required by the model (which by default is set to 0.7), using the argument `--trust_level` explained in the documentation. For example, we chose to predict once again using a trust level of 0.5, obtaining 11 residues of which 5 belonged to the known binding site (as well as the 3 residues predicted using the default trust level). These eleven residues are shown in blue in the following image:

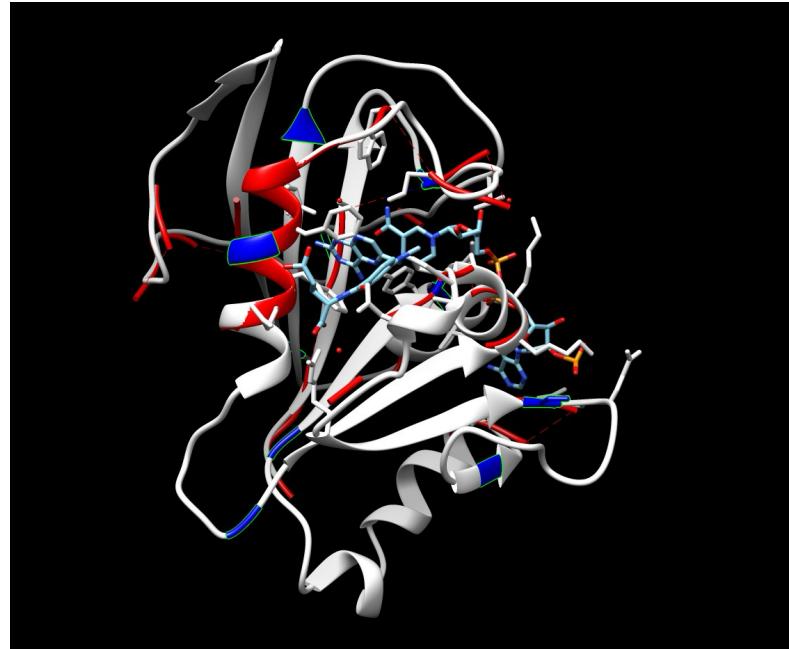


Figure 4. Example 2 with predicted binding sites

Example 3: aromatic amino acid aminotransferase with cyclohexane propionic acid (PDB code: 2AY2)

The *Paracoccus denitrificans* aromatic amino acid aminotransferase is an enzyme involved in the transfer of an amino group (NH_2) from one molecule to another and has contrary properties: flexibility and rigidity (Okamoto et al., 1999). This protein is an homodimer, so this example will help us evaluate how the model performs when there are two or more binding sites. The next image was obtained after superimposing the structures of the protein and its known binding sites, highlighted in red:

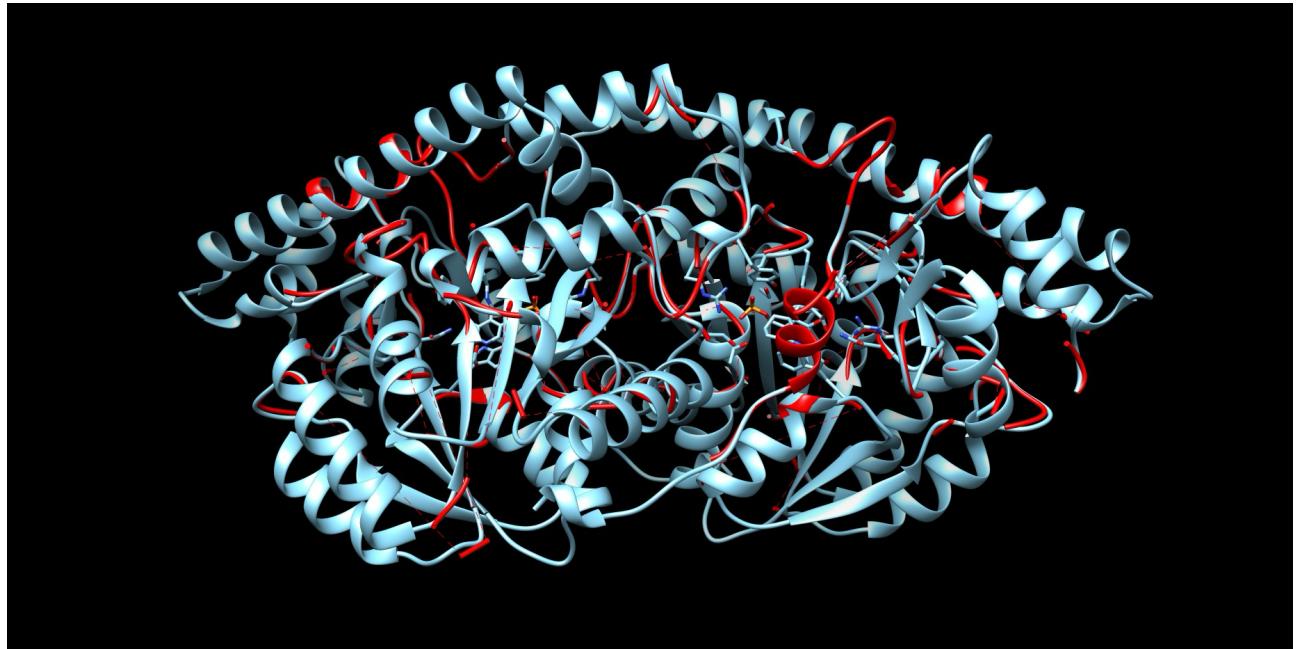


Figure 5. Example 3 with real binding sites

The model predicted 72 residues, 42 of which are found in the known binding site of 223 residues. In the following image we can see the predicted residues highlighted in yellow:

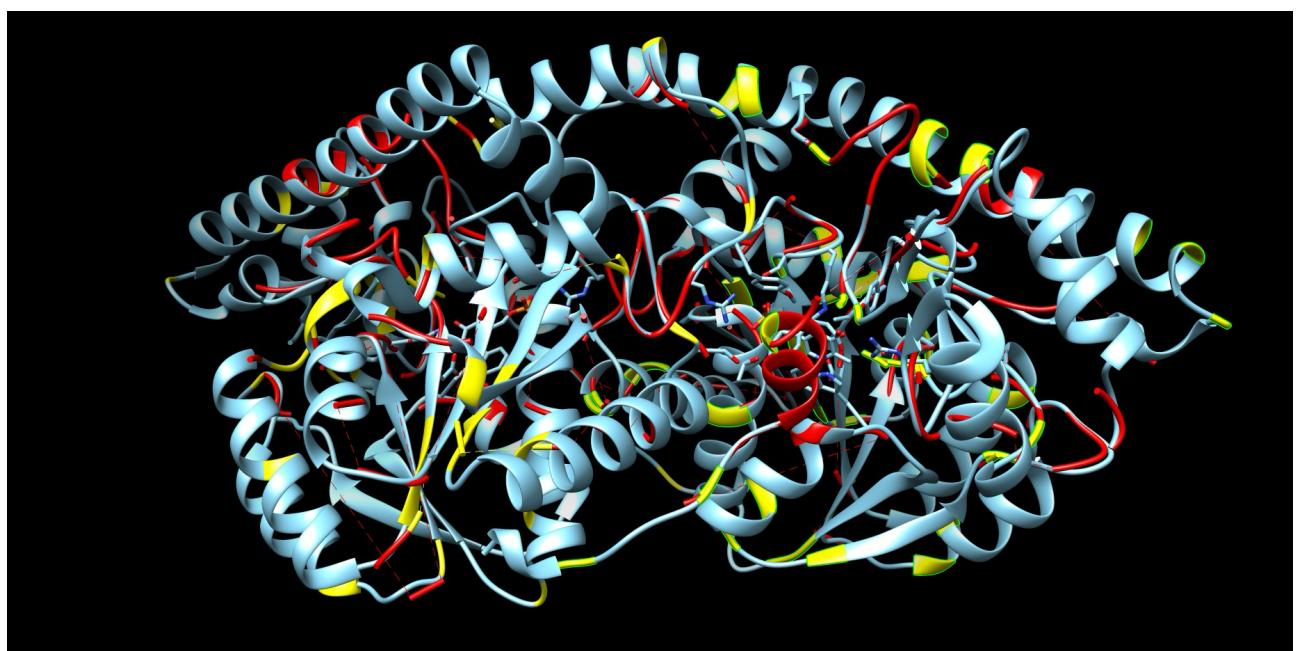


Figure 6. Example 3 with predicted binding sites

As we can see, our model was able to predict residues belonging to both binding sites, with an accuracy similar to the examples shown above.

References:

- Ahmed, A., Mam, B., & Sowdhamini, R. (2021). DEELIG: A Deep Learning Approach to Predict Protein-Ligand Binding Affinity. *Bioinformatics And Biology Insights*, 15, 117793222110303. <https://doi.org/10.1177/11779322211030364>
- Barna, T., Khan, H., Bruce, N. C., Barsukov, I., Scrutton, N. S., & Moody, P. (2001). Crystal structure of pentaerythritol tetranitrate reductase: “flipped” binding geometries for steroid substrates in different redox states of the enzyme. *Journal Of Molecular Biology*, 310(2), 433-447. <https://doi.org/10.1006/jmbi.2001.4779>
- Cody, V., Luft, J., & Pangborn, W. (2005). Understanding the role of Leu22 variants in methotrexate resistance: comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. *Acta Crystallographica Section D: Structural Biology*, 61(2), 147-155. <https://doi.org/10.1107/s0907444904030422>
- Okamoto, A., Ishii, S., Hirotsu, K., & Kagamiyama, H. (1999). The Active Site of Paracoccus denitrificans Aromatic Amino Acid Aminotransferase Has Contrary Properties: Flexibility and Rigidity,. *Biochemistry*, 38(4), 1176-1184. <https://doi.org/10.1021/bi981921d>