LigandMapper.py

ANALYSIS OF EXAMPLES

About the model producibility and evaluation

CONTENT

EXAMPLE ONE | 1HV4.PDB

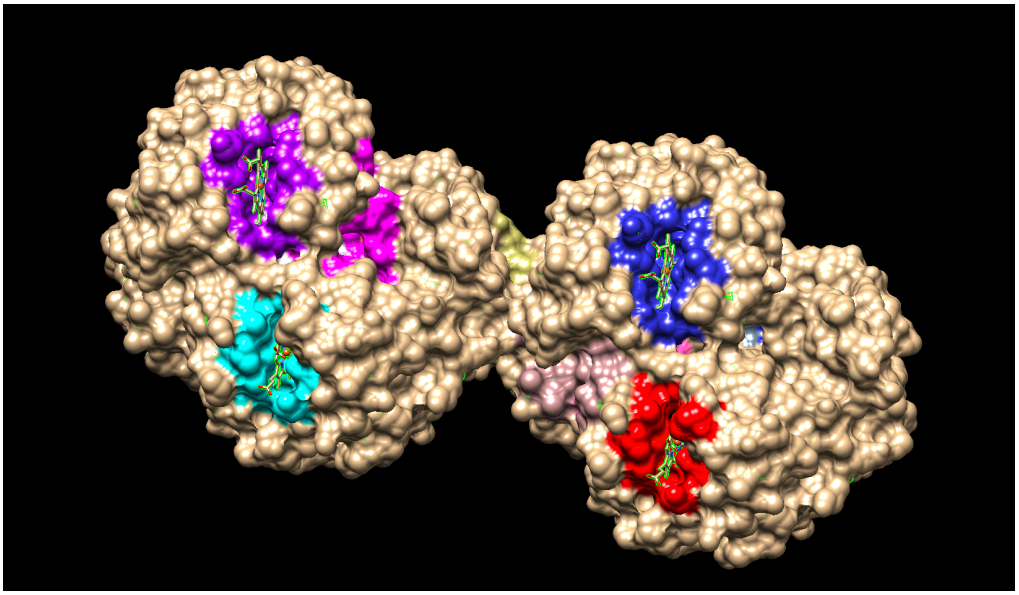
As the first example, to test the model we executed *LigandMapper.py* on **hemoglobin** pdb structure which has been downloaded from the PDB database as *1hv4.pdb*. It has hem groups included in the structure, so the spots where hem binds should be recognized as binding domains. Additionally, we except more putative pockets to be found by the LigandMapper.py.

Based on the generated 1hv4.pdb_predictions.tsv file, our programme predicted 28 pockets with 8 pockets having the probability greater than 0.9. Since the structure that has been put as input is hemoglobin pdb that contains two monomer molecules, these 8 pockets probably correspond to hem-binding sites. When mappings these to their corresponding colours, as each pockets has its own colour based on its rank, it is evident that these are indeed hem-binding sites. A visual representation of the structure with its predicted pockets is given below, observed in *Chimera*.

Pockets are ranked based on their scores which are calculated as the sum of predicted squared positive class probabilities of all inner pocket points, as mentioned in one paragraph before, followed by the probability of finding that particular pocket at a given position. Additionally, as parameters are also given sas_points which correspond to the number of solvent accessible surface points, surf_atoms which is the number of surface atoms and coordinates of a pocket's center given as center_x, center_y and center_z. Other parameters are explained in detail on github repository for the LigandMapper and in the tutorial document file.

name	rank	score	probability	sas_points	surf_atoms	center_x	center_y	center_z
pocket1	1	58.56	986	146	67	63.7172	11.4773	12.4688
pocket2	2	52.37	982	124	56	17.1442	63.1963	53.2115
pocket3	3	51.24	981	128	62	24.1848	45.0528	35.2764
pocket4	4	49.47	979	126	61	33.6694	8.5785	25.9878
pocket5	5	47.50	976	122	58	45.8684	65.3456	36.6562
pocket6	6	46.97	975	140	67	42.5339	-9.8763	9.9624
pocket7	7	44.90	972	116	61	42.9832	20.1710	-0.4937
pocket8	8	40.33	964	124	55	24.7938	75.5823	26.0916
pocket9	9	19.62	842	131	54	43.7281	16.1444	15.5976
pocket10	10	19.43	839	191	87	24.1337	60.7039	31.7572

Output .tsv file from the LigandMapper.py including top 10 pocket predictions for the hemoglobin

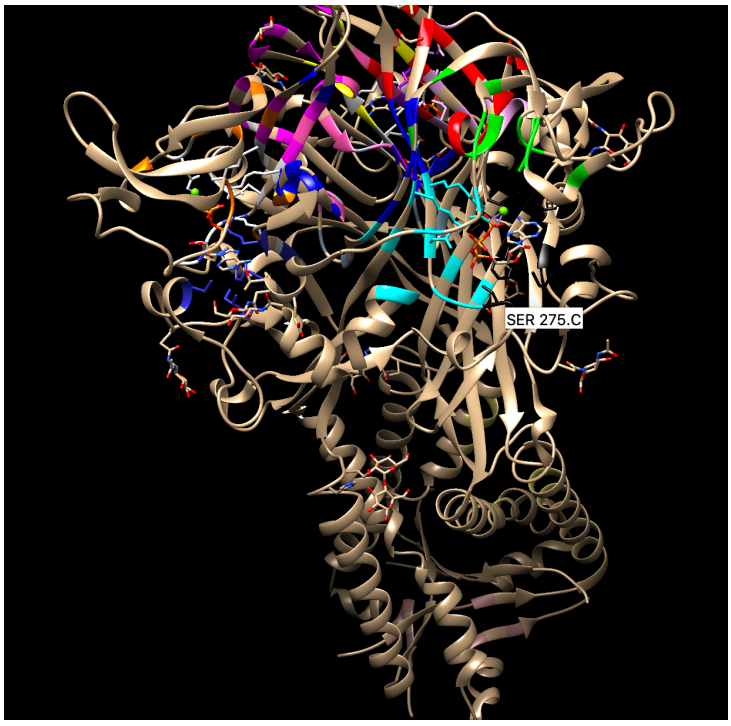


Visual representation of hemoglobin structure with predicted binding sites (Chimera)

EXAMPLE TWO | 6AH5.PDB

Another example that we have decided to include is the ATP-activated **P2X3** receptor expressed in nociceptive sensory neurons. It plays an important role in pain signaling with strong desensitization which depends on the rate of dissociation of the agonist from the binding site. [1] The paper that we used as a reference stated that the fine subunit-specific structural properties predisposing the receptor to tight capture of agonist inside the binding pocket have not been elucidated which is why we wanted to compare the predictions from LigandMapper.py with the ones performed in the scientific paper.

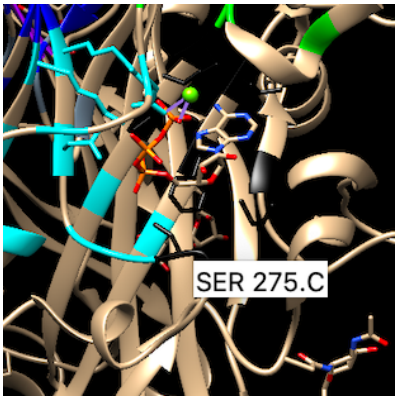
The goal of this example is to predict the binding site within a particular residue S275 which was found to play an important role in receptor desensitization. Additionally, we were interested to find out the rank and the score that this pocket would obtain. PDB structure that was used in this example was downloaded from PDB database as *6ah5.pdb*.



Visual representation of P2X3 receptor structure with predicted binding sites (Chimera)

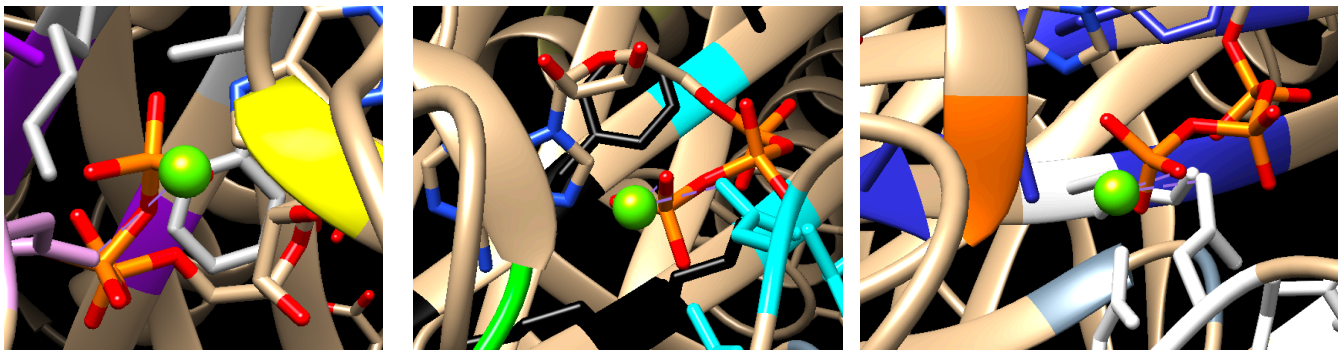
As it can be seen below, even the top one prediction has a lower probability, compared to the predictions of the well-known binding sites of some proteins. S275 has been found incorporated within *pocket5* (coloured in cyan) with rank 5 and the probability of 0.378. Such results suggest that, while S275 plays a role in desensitization of the receptor P2X3, this particular functionality is not amongst the most important ones, since there are pockets that, based on their scores, seem to play a bigger role about the binding properties of this receptor.

name	rank	score	probability	sas_points
pocket1	1	10.20	548	82
pocket2	2	9.09	487	58
pocket3	3	8.61	460	75
pocket4	4	7.63	403	60
pocket5	5	7.26	378	78
pocket6	6	7.06	366	74



Output .tsv file from the LigandMapper.py including top 6 pocket predictions for the P2X3 and the zoom-in of the Serine 275 residue

To find out what are other predictions in this example, we did a further research, since this was not stated in the first referenced paper. We concluded that predictions with the higher score belong to magnesium binding sites. These sites can be seen on the image below.

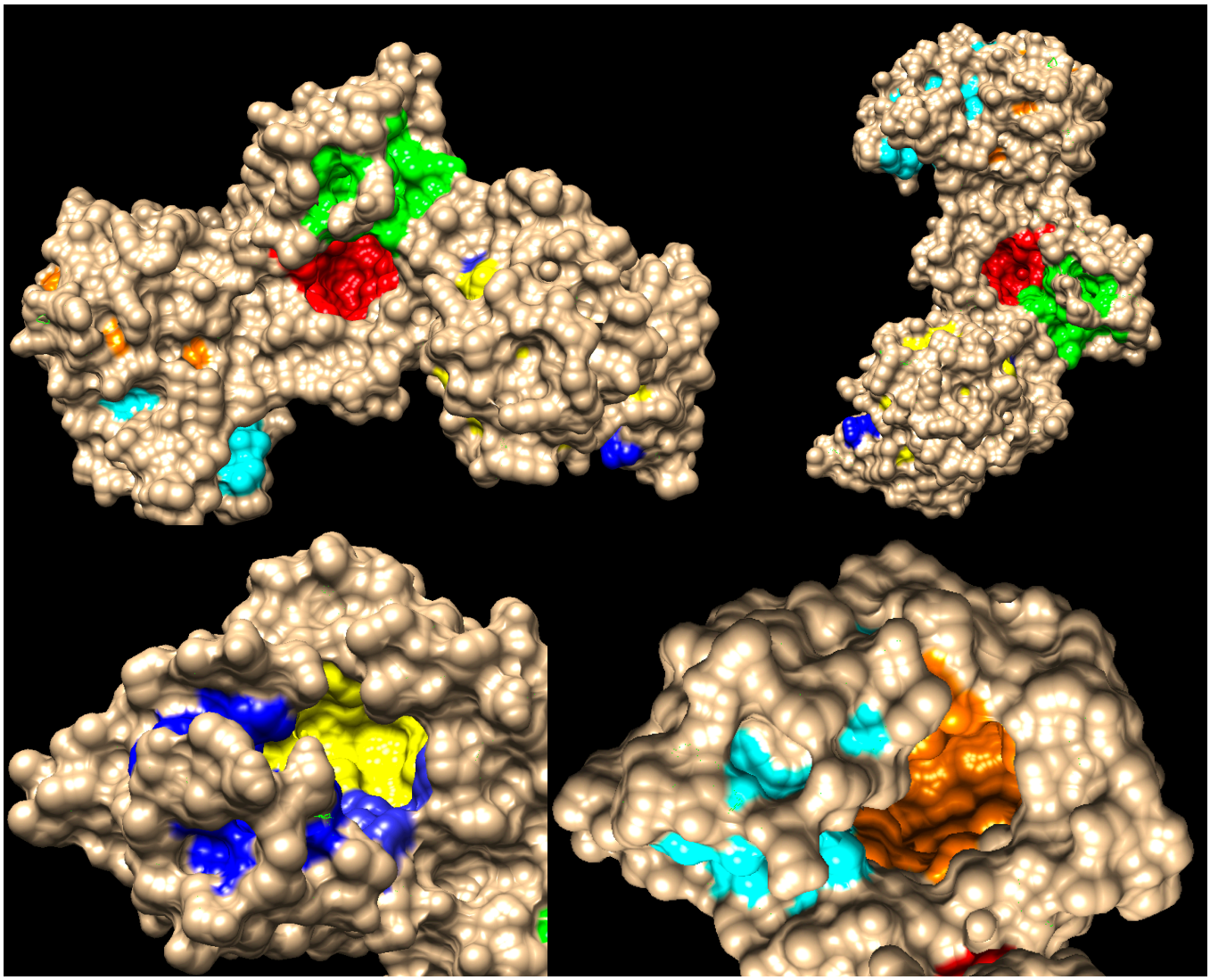


Magnesium binding sites on P2X3 (Chimera)

EXAMPLE THREE | 1G13.PDB

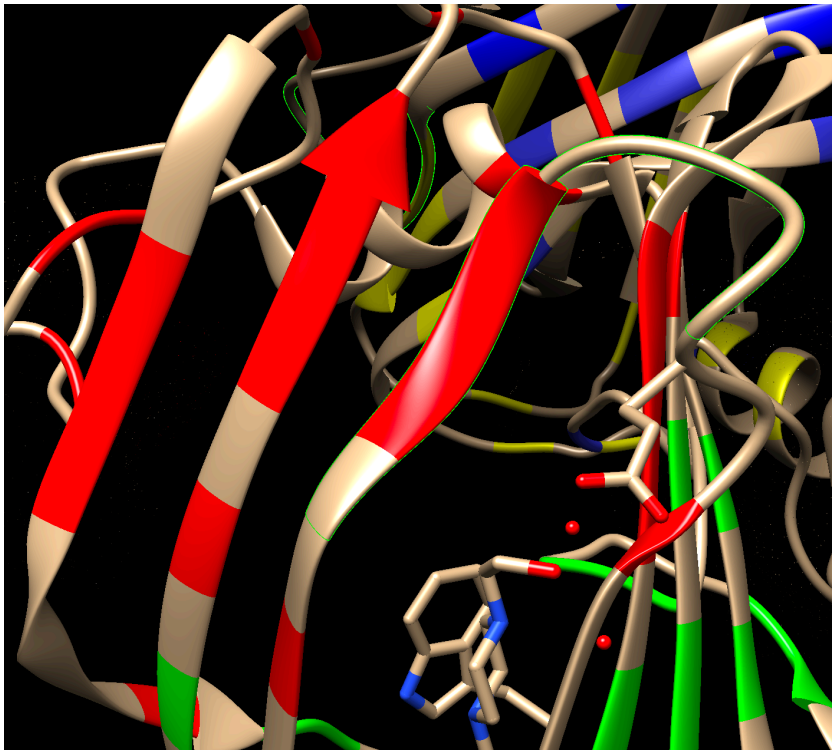
PDB structure *1g13.pdb* belongs to GM2 activator protein (GM2-AP) which belongs to a small group of non-enzymatic lysosomal proteins that act as cofactors in the sequential degradation of gangliosides. One of the more interesting features of the GM2-AP structure is that it possesses an accessible central hydrophobic cavity rather than a buried hydrophobic core. [2] Additionally, it is stated that this cavity is suitable for binding 18-carbon lipid acyl chains, since its dimensions are 12 Å x 14 Å x 22 Å. With this example we wanted to see if, with the prediction with our LigandMapper.py, we are going to get rank 1 for this binding site, since it seems to be quite conserved and important for the functionality of the protein.

By visualizing the output file with the labeled binding sites via Chimera it is evident that there is a cavity in the centre of the protein structure for each chain. On the image below the cavity of one chain can be seen clearly and it is coloured red which corresponds to the first rank. Other two cavities found in the rest two chains are labeled orange and yellow which belong to the second and the third rank, respectively. Therefore, LigandMapper.py successfully recognized GM2 cavities and then sorted them as three most probable ligand binding sites.



Visual representation of GM2 3 chains' structure with predicted binding cavity sites (Chimera)

By looking at a sequence level (searching the part of the sequence in Chimera where the ligand is bind to protein) the binding part is in the red coloured pocket. Sequence where the ligand is making a bond with the protein can be found on PDB database. Thus, it seems once again, that our algorithm predicted correctly the cavity as the most probable to serve as the ligand binding site of GM2-AP.



Ligand binding site located within the red pocket as predicted by LigandMapper.py (visualized in Chimera)

REFERENCES

1. Liu S, Wang M, Wang N, Li S, Sun R, Xing J, Wang Y, Yu S, Li L, Li G, Liang S. Exploring the molecular mechanism of the effect of puerarin on P2X3. Int J Biol Macromol. 2020 Jan 1;142:484-491. doi: 10.1016/j.ijbiomac.2019.09.120. Epub 2019 Oct 5. PMID: 31593721.
2. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. " MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res. 2014 Jan; 42(Database issue):D297-303