

Cheminformatics Final Report

Jan-Erik, Hsin-Hsiang

Overview

In the final project, we took the paper “In silico interaction analysis of cannabinoid receptor interacting protein 1b (CRIP1b) – CB1 cannabinoid receptor” from Pratishtha Singh et al. as a reference example. We learned how to predict the 3D structure of a protein and used in silico method to do protein docking analysis. The final results looked similar to what paper showed. Algorithms for energy minimization and refinement were also included.

Introduction

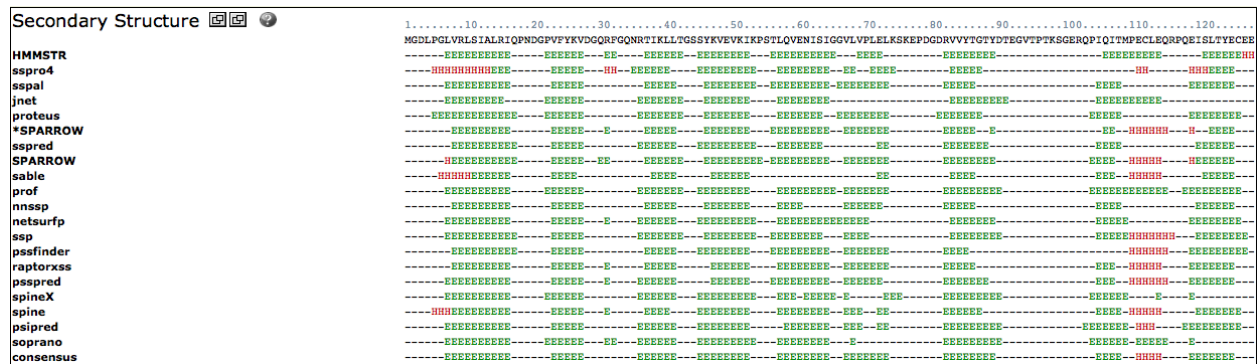
Cannabinoid Receptor (CB₁ receptor) is an important receptor in our body. It regulates the neural system, including appetite, obesity, learning, and neuroprotection. Nie et al. found that the truncation of the distal C-terminal tail of the CB₁ receptor can enhance the constitutive receptor activity (2001). This results encouraged scientists to find proteins that can bind to the distal C-terminal tail of CB₁ receptor because that protein may be a drug candidate that can treat disorders of modulation of CB1 activity.

Niehaus et al. used two-hybrid assay to screen a human brain cDNA library and found that there were two proteins that can bind to the C-terminal tail of CB₁ receptor, which are CRIP1a and CRIP1b. After lots of biological experiments, Niehaus et al. found that the CRIP1a protein can decrease the constitutive activity of CB₁ receptor. However, the function of CRIP1b was still unknown.

Modelling of CRIP1b structure

Secondary structure prediction of CRIP1b

The sequence of CRIP1b was obtained via NCBI as described in the paper. Searching in the database with the keyword “NP_001104571” resulted in a hit for “CB1 cannabinoid receptor-interacting protein 1 isoform CRIP1b [Homo sapiens]”. The sequence of the protein can be accessed by downloading the FASTA format which encodes the amino acids. In order to find a potential template molecule for later modelling and to predict the possible two dimensional features (such as beta sheet or alpha helix) the GeneSilico METASERVER (<https://genesilico.pl/meta2/>) was used. After successfully creating an account the fasta sequence was used as an input for the “Query sequence” and the rest of the options were set to their default values. The results of the secondary structure predictions are combined to form one consensus sequence which is being updated for quite some time after the initial submit. The final consensus sequence can be seen in the following picture.



If we now compare the consensus sequence with the original one proposed by the paper we can see a clear difference in the region of ~110-120bp. In the secondary structure there is a predicted alpha-helix which is not present in the papers prediction. On top of that there are further minor differences but none as significant as the above mentioned one. The algorithms used in our result and the paper are exactly the same nevertheless some of the same algorithms return different results in this exact region.

Tertiary structure prediction of CRIP1b

The tertiary structure predictions of the GeneSilico METASERVER however fail to find a protein that is high enough in similarity to achieve sufficient results. To find a suitable protein the SWISS-MODEL server was utilized where again the Fasta sequence was used as an input for the “Target sequence” and the search for a template was started resulting in 1DS6.B as a suitable molecule with a similarity of 27.5% (the same result as stated in the paper). If we compare the sequence alignment of 1ds6.B and the sequence however our result differs again compared to the reference one.

After downloading the pdb file of 1ds6 from the protein database the previously obtained alignment has to be encoded in a file in order for MODELLER to be able to read it. We used the alignment of secondary structure as presented in the paper and not our own. The alignment saved in a “.ali” file includes both amino acid sequences, the CRIP1b and the 1ds6.b one with the specification of which residues of which subchain of the pdb file of 1ds6 should be used for modelling. In this modelling only the B subchain of 1ds6 was used to do the structure creation. A simple python script is including the MODELLER functions and is defining the secondary structure restraints from the secondary structure prediction mentioned above. Using MODELLER the sequence from CRIP1b is being mapped on the template protein whose three dimensional structure is specified in the according pdb file. An initial model is formed with random positions of molecules satisfying the alignment and restraints which is then optimized using VTFM and conjugate gradients. The structure can be further optimized using molecular dynamic simulations and conjugate gradient refinements. Using the loop model all parts of the protein that are not aligned are modified and optimized. The

standard loopmodel included in MODELLER is described further in Fiser et al., 2000. A more rigorous refinement can also be obtained using the dope loop-modelling which also includes solvent interactions.

The modelling of the 3D structure with MODELLER can be tweaked using different parameters. The total number of models being created, each with a different random starting point can be increased potentially leading to a better result in the end. Similarly the number of loop models created can be changed from one for each model created in the first step to several for each covering a broader range of final configurations. Optimization after the initial model creation can be done either by conjugate gradient or by gc and molecular dynamics which each can be varied in their rigorousness. There also exists an option to include hydrogen atoms present in the pdb-file of the template or present water molecules.

We created different 3d models of the CRIP1b molecule using different settings of the above mentioned parameters. At the end of each modelling run a list of DOPE energy scores of each molecule was created in order to pick the most stable ones. In our modelling runs we were never able to recreate the energy values presented in the paper.

Different runs with different options done:

5 models with each 8 dopeloopmodels with no molecular dynamics, including hydrogen and water molecules: best score: -8716

5 models with each 8 dopeloopmodels with molecular dynamics, including hydrogen and water molecules: best score: -8944

5 models with each 5 loopmodels with no molecular dynamics: best score: -8612

5 models with each 5 loopmodels with molecular dynamics: best score: -8774

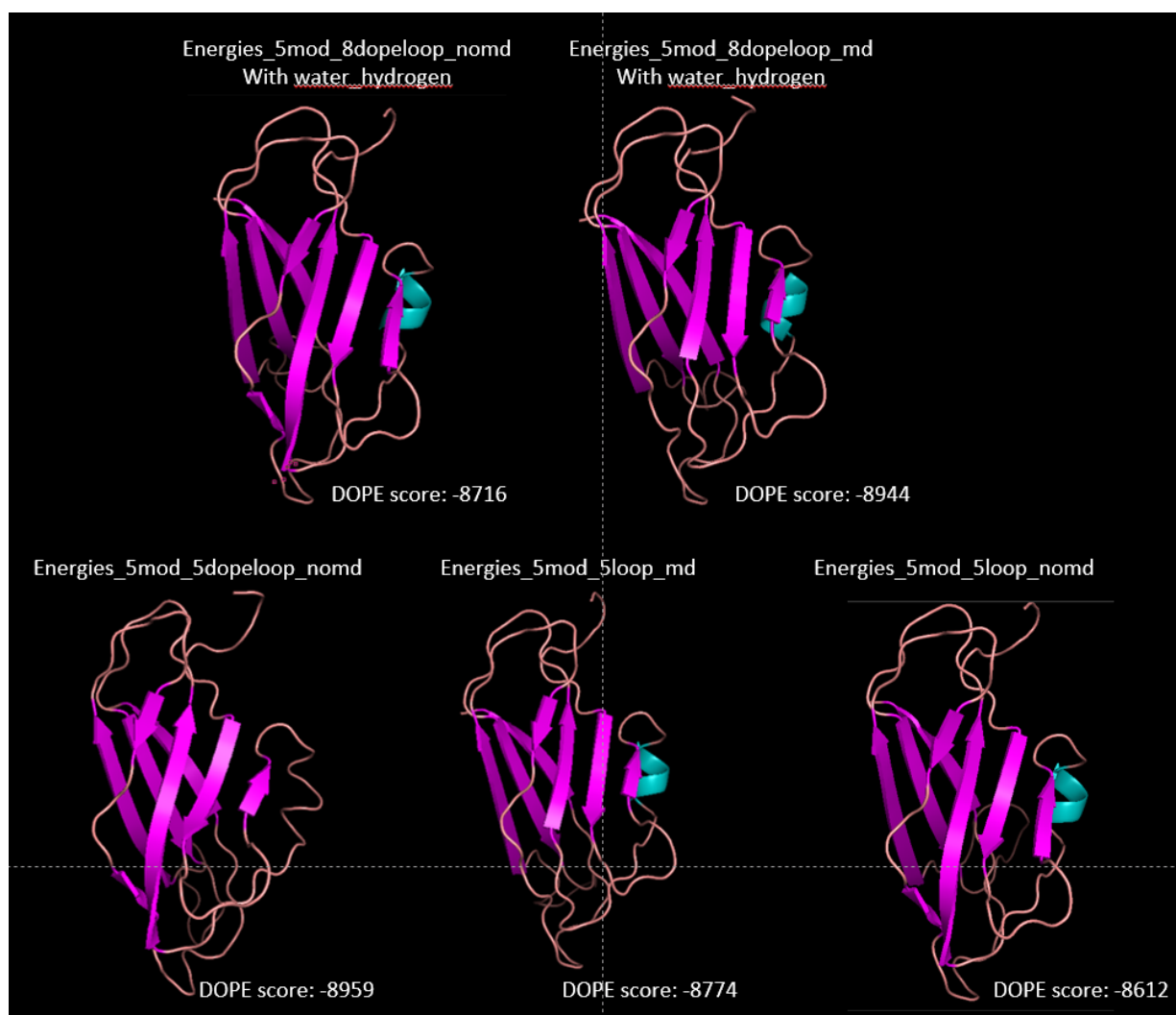
5 models with each 5 dopeloopmodels with no molecular dynamics: best score: -8760

5 models with each 5 dopeloopmodels with no molecular dynamics: best score: -8959

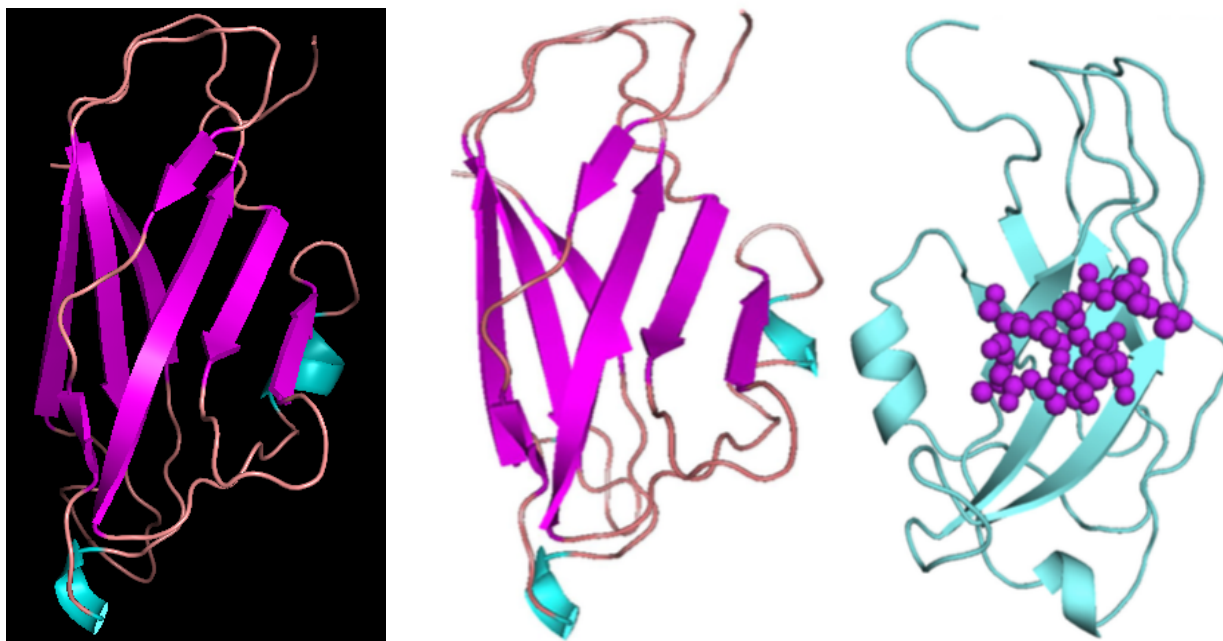
DOPE score presented in the paper: -13222

Even the model with the lowest energy score in our simulation was still ~4000 points below the energy score presented in the paper.

This could be due to different parameters used in the software although a lot of different ones were already tried whereas the paper states no specifics. It might be the case that an increase in the number of models and loop-models created results in finding a more stable model by chance. But the increase in models created always goes along with an increase in time spend calculating those which can become quite time consuming.

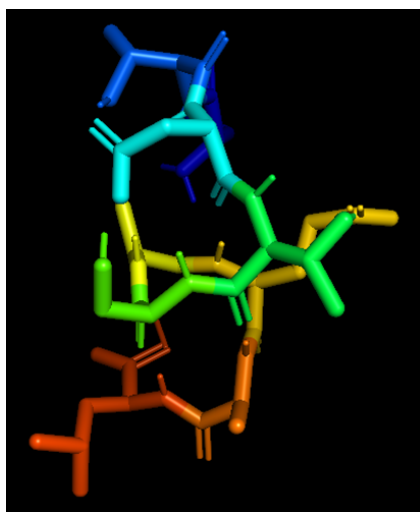


By observing every structure by eye and comparing it to the one proposed in the paper we picked one created with the dope-loop option and molecular dynamics as an optimization step. The left picture is the molecule we created whereas the molecules on the right is the published one. The biggest mismatch visible by eye (between left and middle) is the alpha-helix in the back which is more pronounced in our model. However in the structure on page 319 used by the authors for their docking (picture on the right) the alpha helix looks also bigger, more similar to the one in our model.



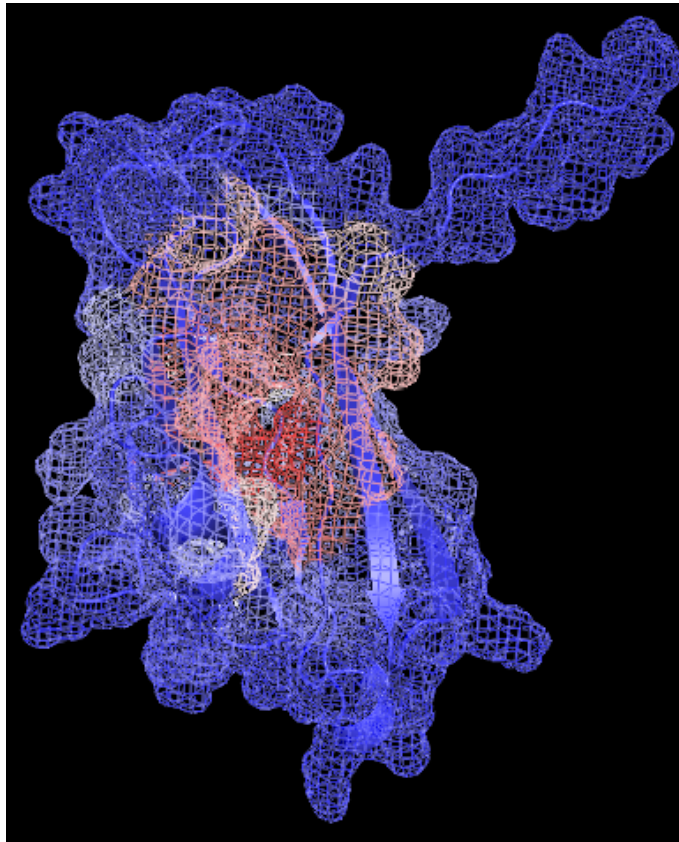
Docking of CRIP1b and CB1

Before the docking can be done the structure of the CB1 peptide has to be known. This is done by inserting the sequence for the CB1 peptide in the PEP-fold server and downloading the pdb file after the run is finished. The obtained structure fits with the one presented in the paper.



The easy interface of the HADDOCK 2.2 server was used to create the docking results. The 3d-structure of CRIP1b selected in the previous step is uploaded and all residues are marked as active, leading to a sampling of the whole protein. The second molecule uploaded to the server is the pdb file of the CB1 peptide where all nine residues are also included as active. After the run finished all results can be

downloaded as a tar file. Using csh scripts all residues can be extracted from each model run where the interaction distance between CRIP1b and CB1 was less than 4 Angstrom. This list contains the amount of times each residue was part of a binding site over all the different docking runs included in the download. If we superimpose those on the pdb file of the molecule we can determine the binding pocket.



By taking the top active residues that take part in the binding of the peptide and protein and comparing it to the ones in the paper we end up with different active residues. This might be due to several reasons. First our initial structure isn't the same as they used in the paper, so a mismatch in active regions is not that surprising. Also the access to the HADDOCK advanced features including the sampling of the docking with water and the tweaking of different parameters is restricted to Expert level status of the account. This is especially important because all active residues reported in the paper bind to the peptide via hydrogen bonding where present water molecules might be a supporting agent.

Secondary/Tertiary structure prediction of CRIP1a

The same procedures were executed using the amino acid sequence of CRIP1a. This molecule was published earlier by Ahmed et al. 2014 however the final three dimensional structure was never published as a downloadable file. One difficulty in

constructing this molecules lies within the fact that the homology against 1ds6.B is not sufficient for the whole sequence. The very end of the sequence has to be mapped using a second template molecule 1r17. The successful modelling depends on the order of those molecules in the alignment file for MODELLER. The resulting molecule has a DOPE score of -12514 which is again lower than the DOPE score originally published in the according paper (-14212). This further increases the reason to believe that some options in the usage of MODELLER are leading to a mismatch in energy scores for CRIP1b and CRIP1a.

