

HW 10

Due Friday, June 12th by 12 noon, emailed to pbradley@fhcrc.org

Simple homology modeling with the BioInfo MetaServer

In this assignment you'll build a homology model and use it to identify residues in a target protein that could potentially interact with a ligand. The target sequence is:

>target

```
MPSGTSQCEDGSAGGFQHLDMHSEKRQLEKGPSGDKDRVWIRPDTPSRCTWQLGRAMADS
PHYHTVLTKSPKILPDILRKIGNTPMVRINKISKNAGLKCELLAKCEFFNAGGSVKDRIS
LRMIEDAERAGNLKPGDTIIEPTSGNTGIGLALAAAVKGYRCIIVMPEKMSMEKVDVLR
LGAEIVRTPTNARFDSPEHVGVAVRLKNEIPNSHILDQYRNASNPLAHYDDTAEIILQQ
CDGKLDMLVASAGTGGTITGIARKLKEKCPGCKIIGVDPEGSILAEPEELNQTEQTAYEV
EGIGYDFIPTVLDRAVVDKWFKSNDEDSFAFARMLIAQEGLLCGGSSGSAMAVAVKAARE
LQEGQRCVVILPDSVRNYMSKFLSDKWMLQKGFMEELS VKRPWWRLRVQELSL SAPLT
VLPVTVCEDTIAILREKGFDAQPVVNESGAILGMVTLGNMLSSLLAGKVRPSDEVCKVLY
KQFKPIHLTDTLGTLSHILEMDHFALVVHEQIQSRDQAWSGVVGPTDCSNGMSSKQQM
VFGVVT AIDLLNFVAAREQTQT
```

Submit this sequence to the Bioinfo MetaServer.

http://meta.bioinfo.pl/submit_wizard.pl

You will get a warning telling you that the same sequence has already been submitted (by me!). Follow the link to the results, which should look something like this:

The screenshot shows the BioInfo MetaServer interface. At the top, the job ID is "Job 26069: cbs mouse [2009-06-05]". The browser address bar shows "http://meta.bioinfo.pl/3djury.pl?meta=v2&id=26069". The page has a green navigation bar with links: "Submit", "Queue", "Servers", "Benchmarks", "PDB-Preview", "Join", "Job 26069: cbs mouse", and "Help".

On the left, there is a sidebar with "FP7 Partner" information and "Protein Modeling Platform" links. The main content area has a search form with "Compute 3D-Jury using:" and "Display:" sections. The "Compute 3D-Jury using:" section has radio buttons for "first", "one", and "all", a "model(s) per server" input, and a "similarity cutoff:" input set to "40". The "Display:" section has buttons for "Submit", "Reset", and "Defaults".

Below the search form is a table titled "3D-Jury - best 20 models". The table has columns: Model, Jscore, Rscore, Fssp, Scop, and PDB Hit. The top hit is RPSB_01 with a Jscore of 257.50 and Rscore of 6e-97. The PDB Hit is "1jbq_A Lyase : Structure Of Human Cystathionine Beta-synthase: A Unique Pyridoxal".

On the right, there is a sequence alignment view. The target sequence is "MPSGTSQCEDGSAGGFQHLDMHSEKRQLEKGPSGDKDRVWIRPDTPSRCTWQLGRAMADS...". The alignment shows a high similarity to the top hit, with a "psipred" line below it. The alignment is shown in a monospaced font with some characters highlighted in blue.

The 3D-Jury method uses consensus between different fold recognition servers to detect distant sequence homology. In this case it's a little silly, since this sequence is very similar to a protein of known structure. The top 3D-Jury hit is to the protein 1JBQ, chain A. If you look at the end of the line corresponding to that hit (it starts with RPSB_01 in the screen shot above; scroll all the way to the right), you'll see a link ([model]) which you can click on to build a homology model using MODELLER (see the figure to the right). You'll need to

```
KQQMVFVVVTAIDLLNFVAAREQTQT
---EEEEEEHHHHHHHHHHHH---[pir]

540 . 550 . 560
KQQMVFVVVTAIDLLNFVAAREQTQT[2D][mqap]
[pir][pdb][model]
[pir][pdb][model]
[pir][pdb][model]
[pir][pdb][model]
[pir][pdb][model]
```

enter your MODELLER license key and an email address to which they'll send the model. The run itself should be pretty fast (an hour or less).

Download the template PDB file 1JBQ from the RCSB PDB website:

<http://www.rcsb.org/pdb/home/home.do>

First we are going to use PyMOL to view our model and generate a new pdb file that contains the model as well as a cofactor, vitamin B6 complex, that is present in the template 1JBQ but not in our model. Once you get the model pdb file by email, open it and the template file in PyMol (see commands below, which assume that the model was saved as "cbs_model.pdb" in the given directory). Notice the long tails that Modeller added outside the region that was covered by the alignment! These cannot be trusted. Also notice that there are two non-protein ligands associated with each chain in 1JBQ: vitamin B6 complex (residue name PLP) and heme (residue name HEM). We want to copy the vitamin B6 from 1JBQ to our homology model, which only contains protein atoms. Note that our MODELLER model won't necessarily align with the chain in the template pdb file (it may have been translated and rotated), so you'll want to use PyMOL's *align* command to put them into the same reference frame. The *align* command is very useful for superimposing structurally similar protein chains. Now you can have PyMOL save a new .pdb file that contains the homology model and the vitamin B6 cofactor. Type the following commands in PyMOL (or load them from a PyMOL script file, see http://pymol.sourceforge.net/newman/user/S0210start_cmds.html), replacing "/Users/pbradley/dat/genome541/HW10/" with the location of cbs_model.pdb and 1jbq.pdb.

```
cd /Users/pbradley/dat/genome541/HW10/  
load cbs_model.pdb  
load 1jbq.pdb  
show cartoon  
hide lines  
select chain_A, 1jbq and chain a  
zoom chain_A  
select vitamin_B6, 1jbq and chain a and resn plp  
select heme, 1jbq and chain a and resn hem  
show sticks, vitamin_B6 or heme  
align cbs_model, chain_A  
save cbs_model_with_vitamin_B6.pdb, cbs_model or vitamin_B6
```

Note that you can type "help <command-name>" within PyMOL to get help on any command. Now modify the script that you wrote for HW9 in order to identify residues in the model whose C-alpha is within 9.0 Angstroms of any atom in the PLP residue.

Finally, accomplish the same thing all within PyMOL. After entering the commands above, try the following:

```
select nearby_C_alphas, ( cbs_model and name ca ) within 9.0 of vitamin_B6  
save nearby_C_alphas.pdb, nearby_C_alphas
```

This should generate a .pdb file with the set of C-alphas that are within 9.0 Angstroms of the vitamin B6. Do the two methods agree?

WHAT TO HAND IN:

1. list of residues in the target sequence with C-alpha atom within 9.0 of vitamin B6.
2. list of residues in the target sequence with C-alpha atom within 9.0 of heme.
3. provide the answer to the following question: why is it necessary to align the model and the template chain? What would happen if we didn't? Would the answers to 1 and 2 be different?