EXAM 1

The exam submission will consist in one compressed directory. In this compressed directory you will include two directories: one named results and another named scratch. In the results directory you will include all the files I ask you during the different questions and a text file where you will answer the questions. The scratch directory will be the directory where you have done the exam, and it will contain all the files you have generated.

Each question accounts for 1 point. Try to keep answers simple and direct. You can use your notes and materials of the classes if you want, you can also check biological databases.

In this exercise we are working with a protein for which we have its sequence in a separate file (target.fa) and its structure in a separate pdb file (target.pdb). Answer the following questions about the protein we are studying:

a) What function do you think it may have? (practicals 1 and 2)

blastp -query target.fa -db /shared/databases/blastdat/pdb_seq -out target_pdb.out

Hemoglobin alpha chain: 1z8u →Hemogglobin's function (search it in uniprot).

Oxygen transport from the lung to the various peripheral tissues.

b) Does it belong to some family of PFAM?. If so, obtain the HMM profile from PFAM and name it p28b.hmm (practical 2)

hmmscan /shared/databases/pfam-3/Pfam-A.hmm target.fa > hmscan.out

hmmfetch /shared/databases/pfam-3/Pfam-A.hmm Globin > p28b.hmm

c) the structure of this protein belongs to some fold in SCOP?, which one? (practical 1)

Search in the SCOP legacy database with the PDB code of the best match of the target sequence → Globin-like

d) Obtain at least three more sequences with known structure and align them with the sequence of the PROBLEM using the HMM profile. Name the alignment p28b.aln (practical 2)

Template PDB ids:
hmmsearch p28b.hmm /shared/databases/blastdat/pdb_seq > p28b_pdb.out

- 1o1n A
- 1abw A
- 1aby A

→ Meto todas en un mismo fasta con la target y lo subo al cluster.
cat target.fa sequences.fa > seqs.fa

hmmalign p28b.hmm seqs.fa > p28b.sto
perl /shared/PERL/aconvertMod2.pl -in h -out c <p28b.sto> p28b.aln

e) Obtain the secondary structure with DSSP and name it p28e.dssp (practical 5)

Extract the secondary structure from the model using DSSP:

mkdssp target.pdb p28e.dssp

f) Do you think the prediction of secondary structure agrees with the actual secondary structure from DSSP?. Obtain an alignment as proof and name it p28f.aln. Is this prediction reliable?(PRACTICAL 5)

1. Extract the sequence of the model with PDBtoSplitChain:

perl /shared/PERL/PDBtoSplitChain.pl -i target.pdb -o target

2. Execute PSIPRED with the model sequence:

runpsipred_single target.fa

3. Transform the model.ss2 file into a pir alignment containing the model sequence and its secondary structure:

perl /shared/PERL/psipred.pl target.ss2 > psipred.pir

4. Extract the secondary structure from the model using DSSP:

mkdssp target.pdb model.dssp

5. Transform the DSSP output into a PIR alignment between the model sequence and its secondary structure.

perl /shared/PERL/aliss2.pl model.dssp > dssp.pir

6. Concatenate both PIR alignments in a single file:

cat psipred.pir dssp.pir > compare.pir

7. Transform the concatenated PIR alignment into a clustalw alignment:

perl /shared/PERL/aconvertMod2.pl -in p -out c <compare.pir>compare.clu

g) Do you think the structure is correct? Can you prove it? show an image of the energies that prove it and name it p28g.png (practical 4)

Prosa.

h) Find if there is some structural problem and show the location with an image (name it p28h.png). What's the sequence fragment with this problem?

i) Do you think the protein PROBLEM can work as a tetramer? show an image that can prove it (name it p28i.png) (practical 3)

1. fetch both PDB (the tetramer and the monomer)
2. remove waters → remove resn hoh
3. rename the tetramer chains (hemoA, hemoB, hemoC and hemoD
4. Now we will superimpose the target monomer on the chain A of the tetrameric 1gli: super target, hemoA
5. Save the target monomer in this new set of coordinates as a new PDB file: save target_A.pdb, target
6. Repeat this same procedure for 1gli chains B, C and D.
7. Open all the target pdb files generated to see the tetrameric form of leghemoglobin

j) Even if it was a monomer, do you think it will work with the function you selected in answer "a"? Use a sequence alignment (ie. you can reuse p28d.aln) to show the conservation/non-conservation of functional residues. Mark the functionally important amino-acids (conserved or non-conserved) in the alignment with the symbol # at the bottom. (practical 2)

k) To further proof your answer in "j": Show an image of the structure of the active site with the relevant aminoacids that preserve (or don't preserve) the function in its active form (i.e. if it requires a cofactor, such as ATP, Ca, etc. include it in the active site interacting with the corresponding residues of the PROBLEM). Name it p28k.png (taking images with pymol)

l) If the structure of the PROBLEM was incorrect (answer in g), try to fix it. Name the new monomer model as p28l.pdb

EXAM 2

The exam submission will consist in one compressed directory. In this compressed directory you will include two directories: one named results and another named scratch. In the results directory you will include all the files I ask you during the different questions and a text file where you will answer the questions. The scratch directory will be the directory where you have done the exam, and it will contain all the files you have generated.

Each question accounts for 1 point. Try to keep answers simple and direct. You can use your notes and materials of the classes if you want, you can also check biological databases.

In this exercise we are working with a protein for which we have its sequence in a separate file (target.fa). Answer the following questions about this protein:

a) Does it belong to some family of PFAM?. If so, obtain the HMM profile from PFAM and name it p28b.hmm (practical 2)
hmmscan /shared/databases/pfam-3/Pfam-A.hmm target.fa > hbtar.out
hmmfetch /shared/databases/pfam-3/Pfam-A.hmm zf-C4 > p28b.hmm

b) Do you think that this protein belongs to some fold in SCOP?, which one? (practical 1)

blastp -query target.fa -db /shared/databases/blastdat/pdb_seq -out target_pdb.out

Template id: 3d24
Search in the SCOP legacy database with the PDB code of the best match of the target sequence → Nuclear receptor ligand-binding domain.

c) What function do you think this protein may have? (practical 1 and 2)
Search 3d24 (template) in uniprot → Steroid hormone receptor.

d) Find 4 homologous proteins in the swissprot database and align them using the HMM you obtained in question a). Then, change the format of this alignment from stockholm to clustalw. Name this alignment homologs.aln. (practical 2)

blastp -query target.fa -db /shared/databases/blastdat/uniprot_sprot.fasta -out target_pdb.out
sp|P11474| sp|Q6QMY5| sp|Q5QJV7| sp|O08580|

hmmalign [markov_model] [multifasta] > name.sto
perl /shared/PERL/aconvertMod2.pl -in h -out c <p28b.sto> p28b.aln

e) Find one template for our protein. (practical 1 and 2)
Blast: 3d24 chain C

f) Use the template you obtained in the previous question to make a model of our target protein. Name this file model.pdb. (practical 4)

Descargar pdb del template y luego:
perl /shared/PERL/PDBtoSplitChain.pl -i template.pdb -o prefix

Practice 4 - Build a model:

1. Find templates: We have done it already with blast.
2. Build a protein model. We need :
    - Target file (target.fa) →yes
    - Script file (modeling.py): contains the executing commands for MODELLER →yes (needs to be modified)
    - Alignment file: contains the alignment between the target and the template/s in PIR format →no

Alignment file:
cat target.fa > target_template.fa
cat template.fasta >> target_template.fa (template found before →download its FASTA)
clustalw2 target_template.fa
perl /shared/PERL/aconvertMod2.pl -in c -out p<target_template.aln>target_template.pir

module load modeller/10.0
source activate modeller
mod10.0 modeling.py

g) Use ProSa to analyze the energies of our model. What structure should you compare your model with? Discuss the energy profile of your model. Save an image of the energy profiles of your model and the protein you decided to compare with as energy_profile.png. (practical 4)

Mirar prosa del model hecho y del pdb de la target y comparar.

h) Compare the secondary structure of your model with the predicted secondary sequence from the sequence. Display this comparison as an alignment including your target sequence, the secondary structure of your model and the predicted secondary structure. Name this file secondary_structure.aln.

1. Extract the sequence of the model with PDBtoSplitChain:

perl /shared/PERL/PDBtoSplitChain.pl -i target.pdb -o target

2. Execute PSIPRED with the model sequence:

runpsipred_single target.fa

3. Transform the model.ss2 file into a pir alignment containing the model sequence and its secondary structure:

perl /shared/PERL/psipred.pl target.ss2 > psipred.pir

4. Extract the secondary structure from the model using DSSP:

mkdssp target.pdb model.dssp

5. Transform the DSSP output into a PIR alignment between the model sequence and its secondary structure.

perl /shared/PERL/aliss2.pl model.dssp > dssp.pir

6. Concatenate both PIR alignments in a single file:

cat psipred.pir dssp.pir > compare.pir

7. Transform the concatenated PIR alignment into a clustalw alignment:

perl /shared/PERL/aconvertMod2.pl -in p -out c <compare.pir>compare.clu

i) This protein can bind ligands such as drugs or hormones. Find an homologous structure that contains one of these ligands. Name it ligand_template.pdb. (practical 3)

Ir mirando templates y ver en el pdb. El que hemos usado anteriormente ya contiene lo que queremos: 3d24

j) Use the structure you obained in the previous question to reconstruct the interaction between our protein of interest and the ligand included in that structure. (practical 3)

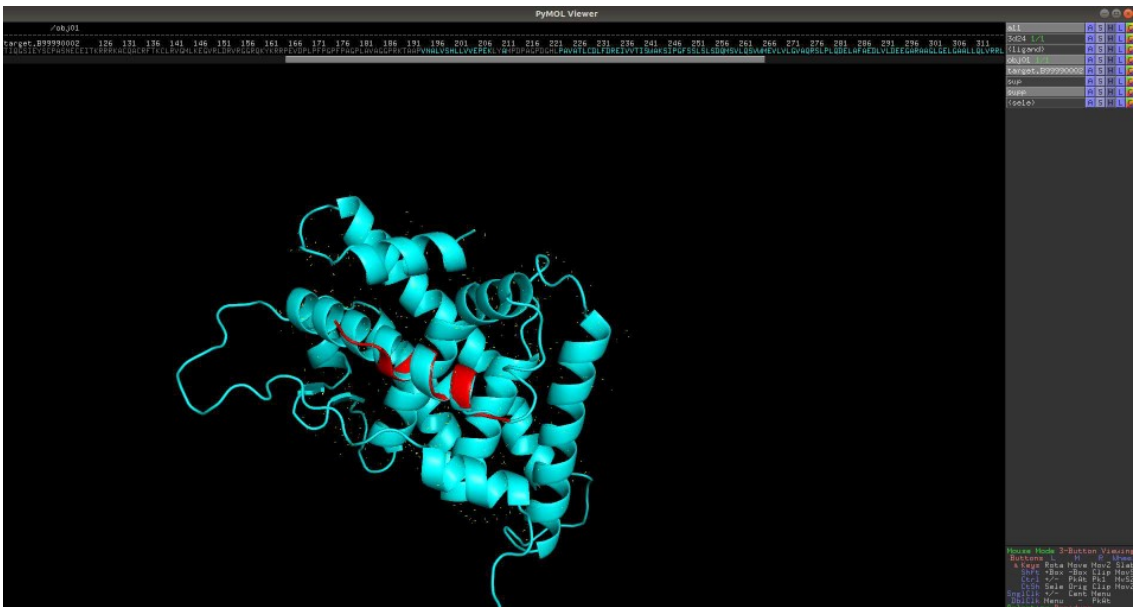
‘’’

https://www.youtube.com/watch?v=mBIMI82JRfI&list=PLUMhYZpMLtaI_Z7to3by2ATHP-cI4ma5X&index=2

show sticks, byres all within 5 of ligand

‘’’

1. fetch 3d24 and model done of our target
2. select ligand - ponerlo en otro color
3. superimponer 3d24 con target



k) Analyze the model of protein-ligand interaction you created in the previous question. Identify the amino acids that are more relevant in the interaction with the ligand. Show these amino acids in images taken with pymol. Name them pymol_image1.png, pymol_image2.png, and so on. (taking images with pymol)

- This will allow us to show residues that could be in the active site: show sticks, byres all within 5 of ligand
- Then we click on the ligand object and: action, find, polar contacts, to any atoms

l) Locate the amino acids you identified in the previous question in the sequence alignment you created in question d. Are these amino acids conserved or not? Explain why you think this is happening? (practical 1 and 2)

The exam submission will consist in one compressed directory. In this compressed directory you will include two directories: one named results and another named scratch. In the results directory you will include all the files I ask you during the different questions and a text file where you will answer the questions. The scratch directory will be the directory where you have done the exam, and it will contain all the files you have generated.

Each question accounts for 1 point, except for questions e and f that count for 2. Try to keep answers simple and direct. You can use your notes and materials of the classes if you want, you can also check biological databases.

In this exercise we will work with two protein structures called SH2.pdb and SH3.pdb. Answer the following questions about these proteins:

a) Obtain the amino acid sequences corresponding with the two structures. Name them SH2.fa and SH3.fa. (practical 4)
perl /shared/PERL/PDBtoSplitChain.pl -i SH2.pdb -o SH2
perl /shared/PERL/PDBtoSplitChain.pl -i SH3.pdb -o SH3

b) Are there HMMs from PFAM for each of these structures? If they are, get them and name them SH2.hmm and SH3.hmm. (practical 2)
hmmscan /shared/databases/pfam-3/Pfam-A.hmm SH2B.fa > hbtar.out
hmmfetch /shared/databases/pfam-3/Pfam-A.hmm SH2 > SH2.hmm

c) Find 4 homologous sequences for SH2. Put them into a single multifasta file. Name the file homologs.fa. (practical 1 and 2)

d) Make a multiple sequence alignment of the 4 sequences you obtained previously using the corresponding HMM you obtained in question c. Then, change the format of the alignment to clustalw format. Name the alignment as MSA.aln. (practical 2)
hmmalign SH2.hmm homologs.fa > homologs.sto
perl /shared/PERL/aconvertMod2.pl -in h -out c <homologs.sto> homologs.aln

e) We want to study the L253R mutation in SH2. Model the structure of SH2 containing this mutation. Be aware that the structure that you have is not complete, therefore the first amino acid corresponds with position 232. Name your model as mutant_model.pdb. (practical 4)

Alignement entre el original y el mutado, target el SH2B.fa mutado manualmente
Practice 4 - Build a model:

1. Find templates: We have done it already with blast.
2. Build a protein model. We need :
    - Target file (target.fa) →yes
    - Script file (modeling.py): contains the executing commands for MODELLER →yes (needs to be modified)
    - Alignment file: contains the alignment between the target and the template/s in PIR format →no

Alignment file:
cat target.fa > target_template.fa
cat template.fasta >> target_template.fa (template found before →download its FASTA)
clustalw2 target_template.fa
perl /shared/PERL/aconvertMod2.pl -in c -out p<target_template.aln>target_template.pir

module load modeller/10.0
source activate modeller
mod10.0 modeling.py

f) Analyze the mutant model you generated previously with ProSa and compare it with a reference structure. Can you tell what is the effect of the mutation in the overall stability of the protein? What reference structure are you using? Take an image of the comparison of the two ProSa profiles and save it as prosa_analysis.png. (practical 4)

g) Search for a structure containing two proteins interacting, one being homolog to SH2 and the other to SH3. Name this file as template.pdb. (practical 1 and 2)

Search in PDB the domains of the structures (we know them from pfam) "SH2, SH3"

h) Use the template you found in the previous question to build a model of the interaction between SH2 and SH3. Name this model interaction_model.pdb. (practical 3)

i) Do you think that the interaction that you modeled in the previous question could happen in nature? Select some amino acids from SH2 that you think are important for the interaction between the two proteins. Show images of these amino acids to proof your answers. Name them aa_picture1.png, aa_picture2.png, and so on. (taking images with pymol)

j) Search for the amino acids that you selected in the previous question in the MSA you created in question d. Are these amino acids conserved? Discuss the degree of conservation of these amino acids and why do you think this is happening. (practicals 1 and 2)

PARTIAL EXAM

In this exercise we are working with a target protein for which we only know its sequence. This sequence is in a separate file (target.fa). All questions are worth one point.

a) What function do you think it may have?

blastp -query targett.fa -db /shared/databases/blastdat/pdb_seq -out target_pdb.out

Cell division protein kinase 2. So we now it's a kinase →attaches a phosphate group to a protein (phosphorylation).

b) Does it belong to some family of PFAM?. If so, obtain the HMM profile from PFAM that fits the best this protein and name it domain.hmm

hmmscan /shared/databases/pfam-3/Pfam-A.hmm target.fa > hbtar.out

hmmfetch /shared/databases/pfam-3/Pfam-A.hmm Pkinase > domain.hmm

c) Does this protein belong to some fold in SCOP?, which one?

De todas las opciones que salgan, escoger la específica de la chain que estamos tratando. (Mirar el uniprot id).

Search the PDB id of the templates in the SCOP database:
http://scop.mrc-lmb.cam.ac.uk/legacy/

If you don't find an entry, you can search in the SCOP2 database:
https://scop.mrc-lmb.cam.ac.uk/

d) Obtain at least three more sequences with known structure and align them with the sequence of the PROBLEM using the HMM profile you obtained in question b. Name this alignment hmm_alignment.aln.

perl /shared/PERL/PDBtoSplitChain.pl -i 1pkd.pdb -o 1pkd

- 1pkd_C
- 1e9h_C
- 1w98_A
- 2jgz_A

hmmalign domain.hmm seqs.fa > alig.sto
perl /shared/PERL/aconvertMod2.pl -in h -out c <alig.sto> hmm_alignment.aln

e) Take the structures for the proteins you aligned in question d and superimpose them. Show the RMSD values that you get for each superimposition. Take a picture of the superimposition and name it superimposition.png.

fetch:
- 1pkd_C
- 1e9h_C
- 1w98_A

remove resn hoh

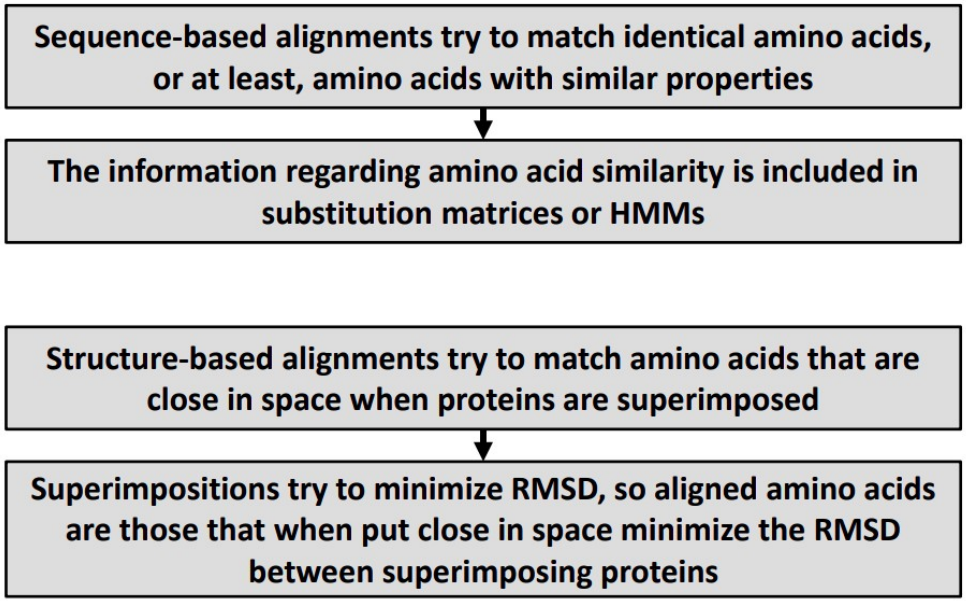
(selecting the corresponding chain and creating objects of these chains)

super 1e9hc, 1pkdc, object=sup4
super 1w98c, 1pkdc, object=sup4

f) Obtain a structure based alignment from the superimposition you made in question e. Name this alignment as structural_alignment.aln.

save structural_alignment.aln, sup4

g) Explain what are the differences between the alignment you obtained in question d (hmm_alignment.aln) and the alignment you obtained in question f (structural_alignment.aln).



h) Use one of the templates you obtained in question d to model the structure of your target protein.

Alignment file:
cat targett.fa > target_template.fa
cat 1pkdC.fa >> target_template.fa (template found before →download its FASTA)
clustalw2 target_template.fa
perl /shared/PERL/aconvertMod2.pl -in c -out p<target_template.aln>target_template.pir

Script file:

```
 alnfile  = 'target_template.pir', # alignment filename
            knowns   = ('1pkdC'),     # codes of the templates
            sequence = 'target')              # code of the target
```

Run:
module load modeller/10.0
source activate modeller
mod10.0 modeling.py

g) PROSA

Since statistical potentials are relative measurements, we should compare the prosa results for our model with the results of a reference protein. Our template is an excellent reference protein because it is an experimental structure (therefore we know it is correct) and it is supposed to be similar to our model. Then, what we have to do is to run two prosa web executions: one for our model and another for our template.

- A Z-score and a plot comparing how good your structure is in comparison with the structures in the PDB.
- An energetic profile showing the statistical potential scores across the amino acid sequence. On the X axis we have the protein sequence, where each amino acid is represented by its position in the protein sequence. On the Y axis we have the energy values for our model. As you see, each amino acid has its own energy value.

- A well modeled region (Green): here the energy values are similar between the model and the template.
- A wrongly modeled region (Red): here the energy values of the model are much higher than the ones of the template. These regions can be corrected, as we will see in the following practice.

EXAM 2020

Practical exam (10p) Reminder: A log of the operations performed (with a brief justification) and answers to questions to be included in p2_answers.txt. We have the sequence of a protein and we wish to know:

a) (0.5p) What should be the function of this protein?

blastp -query target1.fa -db /shared/databases/blastdat/pdb_seq -out target_pdb.out

2p55_A mol:protein length:333 Dual specificity mitogen-activated protein kinase kinase 1

kinase

b) (0.5p) What's the PFAM family domain(s)? Obtain the PFAM domain HMM(s) and name it (them) as p2b.hmmm

hmmscan /shared/databases/pfam-3/Pfam-A.hmm target1.fa > hbtar.out

hmmfetch /shared/databases/pfam-3/Pfam-A.hmm Pkinase> domain.hmm

c) (0.5p) What are the SCOP (or CATH) structural domains?

d) (1.5p) Obtain the structure of this molecule with the side-chains and name it p2d.pdb.
Do you think the function of this protein, described in point "a", can be done?

perl /shared/PERL/PDBtoSplitChain.pl -i 2p55.pdb -o 2p55

Alignment file:
cat target1.fa > target_template.fa
cat 2p55A.fa >> target_template.fa
clustalw2 target_template.fa
perl /shared/PERL/aconvertMod2.pl -in c -out p<target_template.aln>target_template.pir

Script file:
```
 alnfile  = 'target_template.pir', # alignment filename
            knowns   = ('2p55A'),     # codes of the templates
            sequence = '2pA')              # code of the target
```

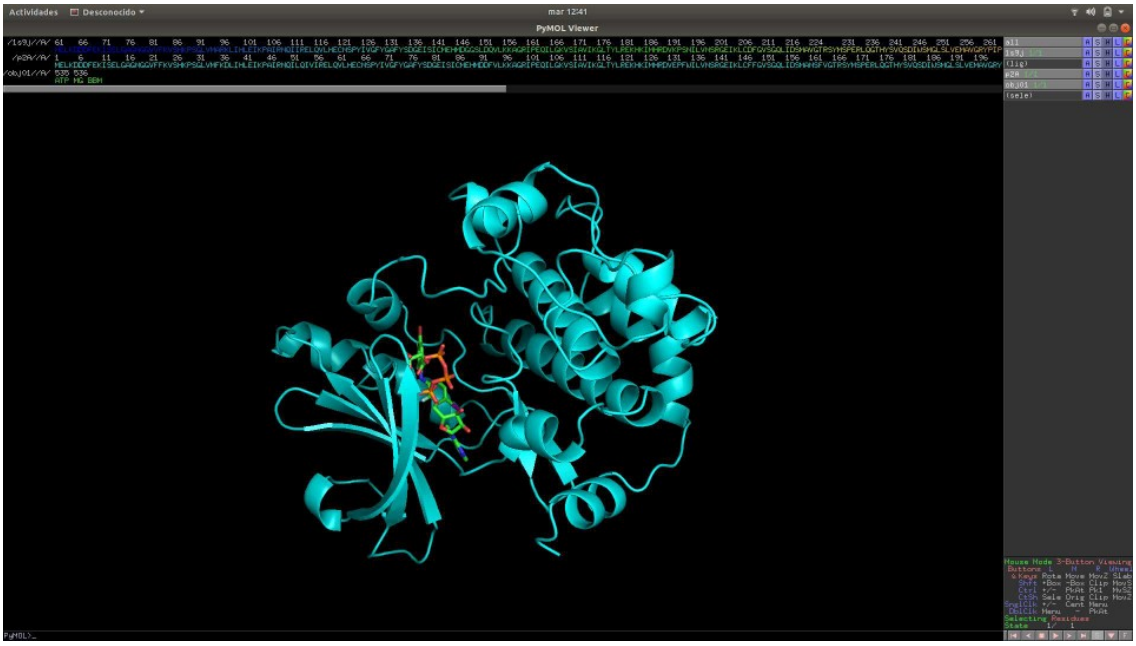
Run:
module load modeller/10.0
source activate modeller
mod10.0 modeling.py

Proof your decision by:

d.1) (1.0p) The comparison of the sequence with other sequences of the same family. Name and mark the important residues in the alignment. Save the alignment in the file p2d1.aln

d.2) (0.5p) Place a ligand molecule relevant for the function of this protein into the model. Name the PDB file with the ligand p2d2.pdb. Name and highlight the residues that should do the function in an image file (p2d2.png)

Fetch 1s9j
Preset>Pretty
Fetch our model
Superimpose: super 1s9j, p2A, object=sup



e) (1.5p) Do you think this structure is OK? Obtain a proof with an image that shows your opinion, locate the problem by identifying the residues in the structure and explain a reason for this problem. Make an image to show the wrong region (p2e.png)

The exam submission will consist in one compressed directory. In this compressed directory you will include two directories: one named results and another named scratch. In the results directory you will include all the files I ask you during the different questions and a text file where you will answer the questions. The scratch directory will be the directory where you have done the exam, and it will contain all the files you have generated.

Each question accounts for 1 point. Try to keep answers simple and direct. You can use your notes and materials of the classes if you want, you can also check biological databases.

In this exercise we are working with a protein for which we have its sequence in a separate file (target.fa) and its structure in a separate pdb file (target.pdb). Answer the following questions about the protein we are studying:

a) What function do you think it may have?

perl /shared/PERL/PDBtoSplitChain.pl -i target.pdb -o target

blastp -query targetA.fa -db /shared/databases/blastdat/pdb_seq -out target_pdb.out

Alkaline phosphatase.

b) Does it belong to some family of PFAM?. If so, obtain the HMM profile from PFAM and name it domain.hmm.

hmmscan /shared/databases/pfam-3/Pfam-A.hmm targetA.fa > hbtar.out

hmmfetch /shared/databases/pfam-3/Pfam-A.hmm Alk_phosphatase > domain.hmm

c) the structure of this protein belongs to some fold in SCOP?, which one?

Alkaline phosphatase-like

d) Obtain at least three more sequences with known structure and align them with the sequence of the PROBLEM using the HMM profile. Name the alignment hmm_alignment.aln.

(cojo las que no tienen e-value 0)
Me descargo sus pdbs, los subo al cluster y hago split chain.

2iuc_A
2glq_A
1zef_A

perl /shared/PERL/PDBtoSplitChain.pl -i 2iuc.pdb -o 2iuc
perl /shared/PERL/PDBtoSplitChain.pl -i 2glq.pdb -o 2glq
perl /shared/PERL/PDBtoSplitChain.pl -i 1zef.pdb -o 1zef

cat targetA.fa 2iucA.fa 2glqA.fa 1zefA.fa > seqs.fa

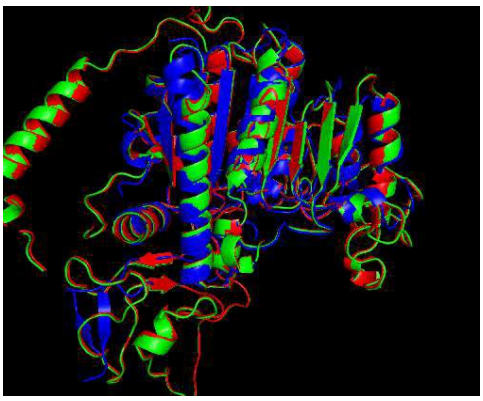
hmmalign domain.hmm seqs.fa > aln_seqs.sto

perl /shared/PERL/aconvertMod2.pl -in h -out c <aln_seqs.sto> aln_seqs.aln

e) Superimpose the three structures and provide an image of the superimposition. Name the image superimposition.png. Then, obtain a structure-based HMM from the superimposition you just obtained. Name this HMM structural.hmm.

(meter en el pymol la chain correspondiente del cluster)
 remove resn hoh

super 1zefA, 2glqA, object=aln
super 2iucA, 2glqA, object=aln



1. save aln.aln, aln →subirlo al cluster

2. Transform from clustalw to stockholm

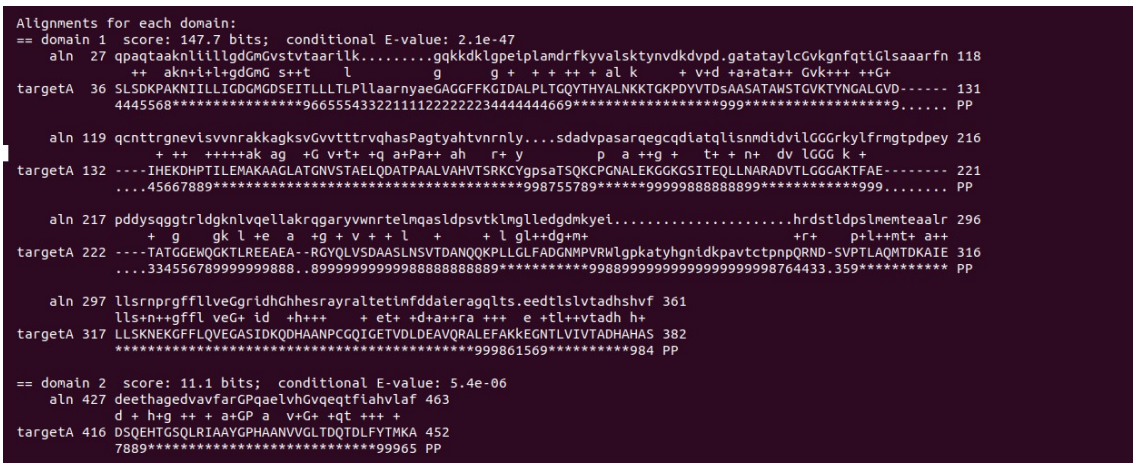
perl /shared/PERL/aconvertMod2.pl -in c -out f <aln.aln> aln.fa
perl /shared/PERL/fasta2sto.pl aln.fa > aln.sto

3. execute hmmbuild to create a HMM

hmmbuild structural.hmm aln.sto

f) Provide an image of the alignment between the HMM you created in the previous question and the sequence of the problem protein. Name the image hmm_match.png. Can you identify what is the region of the protein sequence that is matching the HMM?

hmmsearch structural.hmm targetA.fa > prova.out



g) Obtain the secondary structure with DSSP and name it dssp.dssp.

Alignment file: (alinear con él mismo)

cat targetA.fa > target_template.fa
cat targetA.fa >> target_template.fa
cambiar el nombre a uno de los dos para q sea target (quitar la A)
clustalw2 target_template.fa
perl /shared/PERL/aconvertMod2.pl -in c -out p<target_template.aln>target_template.pir

Script file:

```
alnfile  = 'target_template.pir', # alignment filename
knowns   = ('targetA'),     # codes of the templates
sequence = 'target')              # code of the target
```

Run:
module load modeller/10.0
source activate modeller
mod10.0 modeling.py

cambiar el nombre a un model por target_model.pdb

mkdssp target_model.pdb dssp.dssp

h) Do you think the prediction of secondary structure agrees with the actual secondary structure from DSSP?. Obtain an alignment as proof and name it secondary_structure.aln. Is this prediction reliable?

PSIPRED is a program to predict the secondary structure of one protein. It takes as input one protein sequence in fasta format and returns a prediction of the secondary structure for that sequence.

DSSP is a program that extracts the secondary structure from a PDB file. It takes as input a PDB file and returns, for each amino acid, in which type of secondary structure is placed.

1. Extract the sequence of the model with PDBtoSplitChain: (ya lo hemos hecho)

perl /shared/PERL/PDBtoSplitChain.pl -i target.pdb -o target

2. Execute PSIPRED with the model sequence:

runpsipred_single targetA.fa

3. Transform the model.ss2 file into a pir alignment containing the model sequence and its secondary structure:

perl /shared/PERL/psipred.pl targetA.ss2 > psipred.pir

4. Extract the secondary structure from the model using DSSP: (ya lo hemos hecho)

mkdssp target.pdb model.dssp

5. Transform the DSSP output into a PIR alignment between the model sequence and its secondary structure.

perl /shared/PERL/aliss2.pl model.dssp > dssp.pir

6. Concatenate both PIR alignments in a single file:

cat psipred.pir dssp.pir > compare.pir

7. Transform the concatenated PIR alignment into a clustalw alignment:

perl /shared/PERL/aconvertMod2.pl -in p -out c <compare.pir>compare.clu

---

Looking at the model.horiz file obtained previously. We will only consider reliable regions those with a score of 9.

i) Do you think the structure is correct? Can you prove it? show one or more images of the energies that prove it and name them SP1.png, SP2.png, SP3.png and so on.

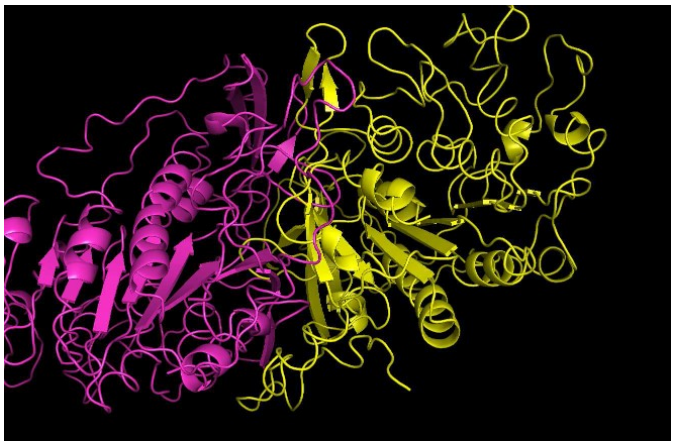
PROSA.

j) Find if there is some structural problem and show the location with an image (name it structural_problem.png). What's the sequence fragment with this problem?

Open the model and the template with pymol and superimpose the two structures.

k) Do you think the protein PROBLEM can work as a dimer? show an image that can prove it (name it dimer.png).

1. Fetch a dimer (homologous from blast are dimers)
2. Select chain1 and chain2
3. Fetch our target model
4.
   super target_model, chain1
   save chainA_tar.pdb, target_model
   super target_model, chain2
   save chainB_tar.pdb, target_model
5. Abrir los nuevos pdbs y ver el dimer



l) Even if it was a monomer, do you think it will work with the function you selected in answer "a"? Use a sequence alignment (ie. you can reuse the alignment from question d) to show the conservation/non-conservation of functional residues. Mark the functionally important amino-acids (conserved or non-conserved) in the alignment with the symbol # at the bottom.

Alkaline phosphatase is a dimer so, if our protein was a monomer it couldn't do the function selected before.

m) To further proof your answer in "j": Show an image of the structure of the active site with the relevant aminoacids that preserve (or don't preserve) the function in its active form (i.e. if it requires a cofactor, such as ATP, Ca, etc. include it in the active site interacting with the corresponding residues of the PROBLEM). Name it active_site.png.

Active site search in uniprot.

Buscar una homologous que tenga ligand y superimponer.

n) If the structure of the PROBLEM was incorrect (answer in question i), try to fix it. Name the new monomer model as corrected_model.pdb.

Mirar el .horiz con valor mayor q 9. Cambiar en el aln.

o) Do you think that your corrected model can work as a dimer? Can you explain why the original model couldn't work as a dimer? Provide pictures of the corrected model and of the original structure to proof your explanations. Name these images corrected_dimer1.png and corrected_dimer2.png.