

my_model

February 20, 2026

1 MY MODEL (P11018)

1.1 1. Find homologs with HMM:

```
psiblast -query P11018.fa -db /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq  
-num_iterations 5 -out P11018_pdb.out
```

And will obtain this:

Query= P11018

Length=319

| Sequences producing significant alignments: | Score (Bits) | E Value |
|--|-----------------|------------|
| 1mee_A mol:protein length:275 MESENTERICOPEPTIDASE | 226 | 2e-72 |
| 1scj_A mol:protein length:275 SUBTILISIN E | 223 | 3e-71 |
| 1yjc_A mol:protein length:275 SUBTILISIN 8397+1 | 218 | 2e-69 |
| 1yjb_A mol:protein length:275 SUBTILISIN 8397+1 | 218 | 2e-69 |
| 1yja_A mol:protein length:275 SUBTILISIN 8397+1 | 218 | 2e-69 |
| 1sbh_A mol:protein length:275 SUBTILISIN 8397+1 | 218 | |

With this I know that 1meeA, 1scjA and 1yjc are the ones I will be using

1.1.1 1.1. Download the fam profile with PFAMa db

Busquem l'HMM de la família utilitzant la base de dades de PFAMa.

```
hmmpfam2 /mnt/NFS_UPF/soft/databases/pfam-2/Pfam_ls P11018.fa > P11018_pfam.out
```

Obtinc:

Scores for sequence family classification (score includes all domains):

| Model | Description | Score | E-value | N |
|--------------|---------------------------------------|--------|----------|---|
| Peptidase_S8 | Subtilase family | 411.9 | 1.1e-120 | 1 |
| LacAB_rpiB | Ribose/Galactose Isomerase | -96.7 | 1 | 1 |
| ETF | Electron transfer flavoprotein domain | -56.0 | 7 | 1 |
| ABC_sub_bind | ABC transporter substrate binding pro | -158.2 | 9.2 | 1 |

Ara sabem que pertanyen a la família de la Peptidase_S8! Ens servirà epr extreure el seu HMM:

With `hmmfetch`, because I already know that P11018 comes from the globin superfamily.

```
hmmfetch /mnt/NFS_UPF/soft/databases/pfam-3/Pfam-A.hmm Peptidase_S8 > Peptidase_S8.hmm
```

1.1.2 1.2. Once have a specific HMM model, use `hmmssearch` against a sequence db (PDB; Uniprot) to find the homologs

```
hmmssearch Peptidase_S8.hmm /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq > Peptidase_S8_pdb.out
```

From this file, I will find the homologous seq. I did it before with psiblast, but the results now are quite different...

So as a templates, I will choose: 1sbhA, 1yjaA and 1yjbA

Un cop triats els templates, anar al fitxer original /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq i copiar les seves seqüències FASTA a un fitxer nou (anomenat templates.fa). To obtain the .fa files for each of this two sequences:

```
cp /mnt/NFS_UPF/soft/databases/pdb/pdb/data/structures/divided/pdb/sb/pdb1sbh.ent.gz .
cp /mnt/NFS_UPF/soft/databases/pdb/pdb/data/structures/divided/pdb/yj/pdb1yja.ent.gz .
cp /mnt/NFS_UPF/soft/databases/pdb/pdb/data/structures/divided/pdb/yj/pdb1yjb.ent.gz .
```

```
gunzip *gz
```

```
cp pdb1sbh.ent 1sbh.pdb
cp pdb1yja.ent 1yja.pdb
cp pdb1yjb.ent 1yjb.pdb
```

```
PDBtoSplitChain.pl -i 1sbh.pdb -o 1sbh
PDBtoSplitChain.pl -i 1yja.pdb -o 1yja
PDBtoSplitChain.pl -i 1yjb.pdb -o 1yjb
```

With the PDBtoSplitChain I have obtained chains A.

```
cat P11018.fa > alignment.fa
cat 1sbhA.fa >> alignment.fa
cat 1yjaA.fa >> alignment.fa
cat 1yjbA.fa >> alignment.fa
```

1.2 2. Align the `hmmfetch` results with the hmm profile using `hmmalign`

Here I want P11018 to align it with other homologs obtained from the `hmmssearch`. A diferència d'un aliniament seqüència-seqüència normal (com faria ClustalW), aquí estic utilitzant una "plantilla intel·ligent" (el model HMM de Pfam) per col·locar cada aminoàcid al seu lloc correcte segons l'evolució de la família de les globines. El programa mira el model HMM, que ja sap quines posicions de la globina són hèlixs, quines són girs i on és probable que hi hagi una inserció o una deleció (gaps). Agafa la teva P11018 i les templates (1abwA, 1o1nA) i les projecta sobre el model una per una. Com que totes s'han alineat contra el mateix motlle, el resultat és que estan alineades entre elles de la manera més precisa possible des d'un punt de vista biològic.

```
hmmalign Peptidase_S8.hmm alignment.fa > final_alignment.sto
```

1.3 3. Convert the sto format into alignment one (clustal)

```
aconvertMod2.pl -in h -out c < final_alignment.sto > final_alignment.clu
```

1.4 3.1. Output of the alignment in clustal format

CLUSTAL W(1.60) multiple sequence alignment

```
P11018 mngeirlipvtneqimdvnelpegi...kvikapemwakgvkgknikvavldtgcdtshpd1  
1abwA -----mlspadktnvkaawgkvga...hageygaealer...mf...sfpttktyfphf  
1o1nA -----mlspadktnvkaawgkvga...hageygaeawerm...mf...sfpttktyfph  
  
P11018 knqiiggknftdddggkedaisdyng...ghthvagtiaandsnggiagvapeasllivkvlg  
1abwA dlshgsaqvkghgkkvadalt...navahvddmpnalsalsdlhahklrvdpvnfkllshcl  
1o1nA fdlshgsaqvkghgkkvadalt...navahvddmpnalsalsdlhahklrvdpvnfkllshcl  
  
P11018 gengsgqyewiinginyaveqkvdiis...msl...lggpsdvpelke-----  
1abwA vtlaahlpaeftpavhasldkflasvstv...tskyrgv...lspaD-----  
1o1nA lvtlaahlpaeftpavhasldkflasvstv...tskyrv...lspaD-----  
  
P11018 -----A-----  
1abwA -----  
1o1nA -----  
  
P11018 -----vknavkngvlvvcaagnegdg...derteelsyp  
1abwA -----ktnvkaawgkvga...hageygaealer...mf...sf  
1o1nA -----ktnvkaawgkvga...hageygaeawerm...mf...sf  
  
P11018 aayneviavgsvsvarelsefsnankeidl...vapgenil...stlpnkkygkltgt...smaaphvs  
1abwA ttkt...fphfdlshgsaqvkghgkkvadalt...navahvddmpnalsalsdlhahklrvdpvn  
1o1nA ttkt...fphfdlshgsaqvkghgkkvadalt...navahvddmpnalsalsdlhahklrvdpvn  
  
P11018 galaliksye...esfqrklsesevfaqlirrt...pldiaktlagngflyltapdelaekaeq  
1abwA fkllshcllv...tlaahlpaeftpavhasldkflasvstv...sky...r-----  
1o1nA fkllshcllv...tlaahlpaeftpavhasldkflasvstv...sky...r-----  
  
P11018 shlltl  
1abwA -----  
1o1nA -----
```

1.5 4. MODELLER

Ara fem servir el programa MODELLER per construir el nostre model. Modeller és una implementació automàtica d'una aproximació de modelatge comparatiu que segueix restriccions espacials. Funciona de tal manera que primer, les estructures 3D de les templates s'alinien amb la target, després s'apliquen les restriccions espacials (distància entre Calpha, cadenes laterals, angles...) que es transfereixen de les templates a la target, finalment, s'obté el model 3D. Aquest complirà al màxim de restriccions possibles.

Modeller només llegirà l’alineament si aquest està en format pir. El programa `aconvertMod2` permet fer aquest canvi de format. L’output és un alineament amb tants blocs com seq i cada bloc té un encapçalament i dos línies amb etiquetes. Aquestes etiquetes hauràn de ser les mateixes que les que posarem a l’input file.

```
aconvertMod2.pl -in c -out p < final_alignment.clu > final_alignment.pir
```

1.5.1 4.1. Modify `modeling.py` file

Aquest file conté els paràmetres que necessita Modeller i les opcions que volem fer servir per a construir el meu model 3D.

- `alnfile` on es posa el nom de l’alinhimanet en format pir, a partir del qual es construeix els models. Entre cometes!
- `knowns` on es posa el nom dels templates utilitzats. Han de tenir el mateix nom que en l’alinhiament pir. Entre cometes i separat entre cometes!
- `sequence` nom del target (en aquest cas P11018) que també ha d’apareixer a l’alinhiament pir.

Els pdb han d’estar al mateix directori

- `a.ending_model = 2` perquè es vol dos models, tot i que es podria fer tants models com vulgui perquè hi ha més d’una solució per a cada restricció espacial.
- `a.make()` NO ES TOCA! Serveix per fer homology modeling de la manera més simple.

1.6 4.2. Run `modeling.py`

Un cop canviat els paràmetres, correm el file .py:

```
modpy.sh python3 modeling.py
```

Com a resultat, obtindré dos models:

```
>> Summary of successfully produced models:  
Filename          molpdf  
-----  
P11018.B99990001.pdb      7236.42578  
P11018.B99990002.pdb      7223.64941
```

Per més comoditat, els canviaré el nom:

```
mv P11018.B99990001.pdb model1.pdb  
mv P11018.B99990002.pdb model2.pdb
```

1.7 4.3. Visualization on chimera

Visualització amb chimera!!

```
chimera model1.pdb model2.pdb 1sbhA.pdb 1yjaA.pdb 1yjbA.pdb
```

I superimosem. Es veu que no acaba de coincidir massa i això pot ser degut a que hi ha gaps al principi i/o al final de l’alinhiament. Llavors es recomanaable construir nous models sense aquest segment, esborrant certs aminaocids de l’arxiu fasta.

1.8 4.4. Modify alignment to enhance the superimposition

EL ALINIAMENT SEGÜENT NO CORRESPON AL ALINIAMENT ORIGINAL, EM VAIG EQUIVOCAR I NOMÉS MOSTRO UN EXEMPLE PER ACLARIR EL QUE ES VOL FER!! Abans:

CLUSTAL W(1.60) multiple sequence alignment

P11018 mngeirlipyvtneqimdvnelpegikvikapemwakgvkgknikvavldtgcdtshpd1
1abwA -----mlspadktnvkaawgkvgahageygaealerflsfpttktyfphf
1o1nA -----mlspadktnvkaawgkvgahageygaeawermflsfpttktyfph

P11018 knqiiggknftdddggkedaisdyngthvagtiaandsnggiagvapeasllivkvlg
1abwA dlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvnfkllshcl
1o1nA fdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvnfkllshcl

P11018 gengsgqyewiinginyaveqkvdiismslggpsdvpelke-----
1abwA vtlaahlpaeftpavhasldkflasvstvltskyrgvlsplD-----
1o1nA lvtlaahlpaeftpavhasldkflasvstvltskyrvlsplD-----

P11018 -----A-----
1abwA -----
1o1nA -----

P11018 -----vknavkngvlvvcaagneggderteelsyp
1abwA -----ktnvkaawgkvgahageygaealerflsf
1o1nA -----ktnvkaawgkvgahageygaeawermflsf

P11018 aayneviavgsvsvarelsefsnankeidlwapgenilstlpnkkygkltgtsmaaphvs
1abwA ttktfyphfdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvn
1o1nA ttktfyphfdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvn

P11018 galaliksyeeesfqrklsesevfaqlirrtlpldiaktlagngflyltapdelaekaeq
1abwA fkllshcllvblaahlpaeftpavhasldkflasvstvltskyr-----
1o1nA fkllshcllvblaahlpaeftpavhasldkflasvstvltskyr-----

P11018 shlltl
1abwA -----
1o1nA -----

Ara:

CLUSTAL W(1.60) multiple sequence alignment

P11018 qimdvnelpegikvikapemwakgvkgknikvavldtgcdtshpd1
1abwA mlspadktnvkaawgkvgahageygaealerflsfpttktyfphf
1o1nA mlspadktnvkaawgkvgahageygaeawermflsfpttktyfph

P11018 knqiiggknftdddggkedaisdyngthvagtiaandsnggiagvapeasllivkvlg

1abwA dlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvnfkllshcl
1o1nA fdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvnfkllshcl

P11018 gengsgqyewiinginyaveqkvdiisimslggpsdvpelke-----
1abwA vtlaahlpaeftpavhasldkflasvstvltskyrgvlsplaD-----
1o1nA lvtlaahlpaeftpavhasldkflasvstvltskyrvlsplaD-----

P11018 -----A-----
1abwA -----
1o1nA -----

P11018 -----vknavkngvlvvcaagneggderteelsyp
1abwA -----ktnvkaawgkvgahageygaealerflsf
1o1nA -----ktnvkaawgkvgahageygaeawermflsf

P11018 aayneviavgsvsvarelsefsnankeidlwapgenilstlpnkkygkltgttsmaaphvs
1abwA ttktfyphfdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvn
1o1nA ttktfyphfdlshgsaqvkghgkkvaltnavahvddmpnalsalsdlhahklrvdpvn

P11018 galaliksyeesfqrlsesevfaqlirrtlpldiaktlagng
1abwA fkllshcllvlaahlpaeftpavhasldkflasvstvltskyr
1o1nA fkllshcllvlaahlpaeftpavhasldkflasvstvltskyr

Ara, tornem a fer:

```
aconvertMod2.pl -in c -out p < final_alignment.clu > final_alignment.pir  
modpy.sh python3 modeling.py
```

```
mv P11018.B99990001.pdb model1_withouttail.pdb  
mv P11018.B99990002.pdb model2_withouttail.pdb
```

```
chimera model1_withouttail.pdb model2_withouttail.pdb 1sbhA.pdb 1yjaA.pdb 1yjbA.pdb
```