PRACTICAL 1

Executing BLAST with our target sequence in the PDB

- Comparison tool used for identifying homologous proteins in the PDB database

  blastp -query TARGET.fa -db ~/Documents/databases/pdb_seq -out TARGET_PDB.out

  Also, to know what is the function of the protein that we found

- Use UniProt database to create an accurate PSSM (unbiased and non-redundant DB)

  psiblast -query TARGET.fa -num_iterations 5 -out_pssm TARGET.pssm -out TARGET.out -db ~/Documents/databases/uniprot_sprot.fasta

  TARGET.pssm → where the position-specific scoring matrix (PSSM) generated during the PSI-BLAST iterations will be saved.

  TARGET.out →where the results of the PSI-BLAST search will be saved.
  Results per iteration are indicated by the word "**Round**"

- Use this accurate PSSM to search for templates in the PDB

  psiblast -db ~/Documents/databases/pdb_seq -in_pssm TARGET.pssm -out RESULT.out

Create our PSSM

- Get a set of sequences from uniprot

perl ~/Documents/perl_scripts/FetchFasta.pl -i file.list -d ~/Documents/databases/uniprot_sprot.fasta -o file.fasta

DO NOT FORGET TO INCLUDE THE TARGET SEQUENCE

- Put them in a MSA with the target using clustalw

  cat TARGET.fa > pssm.fasta
  cat file.fasta >> pssm.fasta
  clustalw2 pssm.fasta (run clustalw)

  Then, change the format of the alignment to fasta format:

  perl ~/Documents/perl_scripts/aconvertMod2.pl -in c -out f <pssm.aln>pssm.fa

- Input the generated MSA into psiblast

  psiblast -in_msa pssm.fa -out target_pdb_specific.out -db ~/Documents/databases/pdb_seq

PRACTICAL 2

- Create a HMM from a MSA using hmmbuild

  hmmbuild MODEL.hmm ALIGNMENT1.sto

- Search for templates using hmmsearch or similar sequences

  hmmsearch MODEL.hmm ~/Documents/databases/pdb_seq > RESULT.out

  - We can also search a domain within a single sequence:

    hmmbuild MODEL2.hmm ALIGNMENT2.sto
    hmmsearch MODEL2.hmm ONE_SEQUENCE.fa > RESULTS2.out

- Using HMM databases (PFAM)
  Searching for the best HMM in a database, for a given sequence.

  (concatenate all the generated HMMs in one file) **cat MODEL1.hmm MODEL2.hmm MODEL3.hmm > minifam**

- Join some HMM and then:
  hmmpress minifam

This is our database, but we can use another one like PFAM. Assign the best profile(s) to the target sequence using hmmscan:

hmmscan minifam TARGET_SEQ.FA > OUTPUT.out
hmmscan /shared/databases/pfam-3/Pfam-A.hmm TARGET_SEQ.FA > PROFILES.out

Inspect the output file and keep the best “Model” because we will extract it using:

hmmfetch ~/Documents/databases/Pfam-A.hmm [name_HMM] > [file_HMM]

[name_HMM] → for example “domain_hbb”