Each state produces amino acids with different frequencies. Then, HMMs change from one state to another while they are producing the protein sequence. In particular, we will work with HMM which contain the following states:

Emission: conserved position of the sequence in reference with the ancestral sequences contained in the HMM.

Insertion: insertion of an amino acid in the sequence in reference with the ancestral sequences contained in the HMM.

Deletion: elimination of one amino acid in the sequence in reference with the ancestral sequences contained in the HMM.

hmmbuild [model_HMM] [alignment]

```
hmmbuild  globins4.hmm  globins4.sto
```

For alignments we will always use Stockholm format.

globins4.hmm: This is the output file name where the generated HMM will be saved.

We can open the file globins4.hmm to check each column ad row. Each position has the logarithm of the probability of emission of a residue. The Aa order is defined in two specific rows of the header (HMM). For each position we have probabilities on two different states (insertion and main), then we have a third row per position with the probabilities of transitions.

Sequence search with HMM is based on finding sequences that fit with the model (finding those sequences that would be generated by the HMM with high probability). This can be done using hmmsearch

hmmsearch [model_HMM] [database] > [output]

```
hmmsearch  globins4.hmm  /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq
                                  >  globins_pdb.out
```

Our outputs will also be classified with e-values.

CONCATANATE HMM

cat globins4.hmm fn3.hmm Pkinase.hmm > minifam (output file)

To check sequences and profiles very fast, we compress and index the database file using **hmmpress**

hmmpress minifam

```
hmmscan  minifam  7LESS_DROME.fa  >  7LESS_DROME_minifam.out
```

Performing multiple sequence alignments using hmmalign

```
hmmalign globins4.hmm globins45.fa > globins45_hmm.sto
```

```
clustalw globins45.fa
```

As hmm works with Stockholm extension and clustalw has it's own we have a code in order to pass from one format to another.

```
perl /mnt/NFS_UPF/soft/perl-lib/aconvertMod2.pl -in h -out c
<globins45_hmm.sto>globins45_hmm.clu
```

HOMOLEG SEQUENCES WITH HMM

phmmer [target_sequence] [database_of_sequences] > [output]

jackhmmer [target_sequence] [database_of_sequences] > [output]

The difference between phmmer and jackhmmer is that jackhmmer perform several iterations (up to a maximum of five by default, which can be changed using the –N flag) and it generates internally a HMMER profile at each iteration, while phmmer is like jackhmmer at iteration 1.

```
hmmscan /mnt/NFS_UPF/soft/databases/pfam-3/Pfam-A.hmm hbb_human.fa
> hb_human_db.out
```

Assign the best profile(s) to the target sequence (hbb_human) using hmmscan.

We need to know the HMM of the profile(s) assigned to our target sequence. They can be extracted from the PFAM database using the program hmmfetch.

hmmfetch [database_HMM] [name_HMM] > [file_HMM]

Search for sequences with known structure that contain the same domain as our target using hmmsearch:

```
hmmsearch domain_hbb.hmm /mnt/NFS_UPF/soft/databases/blastdat/pdb_seq
> hbb_pdb_by_HMM.out
```